DOCUMENT RESUME

ED 416 254 TM 028 134

AUTHOR	Green, Patricia J.; Hoogstra, Lisa A.; Ingels, Steven J.; Greene, Harrison N.; Marnell, Patricia K.
TITLE	Formulating a Design for the ECLS: A Review of Longitudinal Studies. Working Papers Series.
INSTITUTION	National Opinion Research Center, Chicago, IL.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NCES-WP-97-24
PUB DATE	1997-08-00
NOTE	125p.
CONTRACT	RN940940001
AVAILABLE FROM	U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 555 New Jersey Avenue, N.W., Room 400, Washington, DC 20208-5654.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Children; Cognitive Tests; Educational Assessment; Educational Research; *Elementary Education; Foreign Countries; Literature Reviews; *Longitudinal Studies; *National Surveys; *Research Design; Research Methodology; *Sampling
IDENTIFIERS	Early Childhood Longitudinal Study; National Education Longitudinal Study 1988

ABSTRACT

The Early Childhood Longitudinal Study (ECLS) is a new study that will focus on children's early school experiences beginning in kindergarten. Approximately 23,000 children will be selected as they enter kindergarten and followed through fifth grade. Base-year data will be collected in the fall of 1998, but there will a field test in the 1996-97 school year. This paper, prepared in support of the development of the ECLS, reviews nine studies, each of which may provide some design features that would be useful in the ECLS. The studies reviewed are the: (1) Beginning School Study; (2) Children of the National Longitudinal Study of Youth; (3) Greensboro Early Schooling Study; (4) Prospects: The Congressionally Mandated Study of Educational Growth and Opportunity; (5) District of Columbia Early Learning and Early Identification Study; (6) National Education Longitudinal Study of 1988 (NELS:88); (7) Canadian National Longitudinal Survey of Children; (8) National Survey of Children; and (9) National Child Development Study. In reviewing the design and results of these studies, several cross-cutting issues were recognized, including the cognitive assessments, the social and emotional measurements, and the measures of environment. Issues concerning overall study design are also explored, and individual summaries are provided for each of the studies. Study design and sampling and administration procedures will be largely based on the experience of the NELS:88. Each of the study summaries contains references. (Contains six tables.) (SLD)

ED 416 254

NATIONAL CENTER FOR EDUCATION STATISTICS

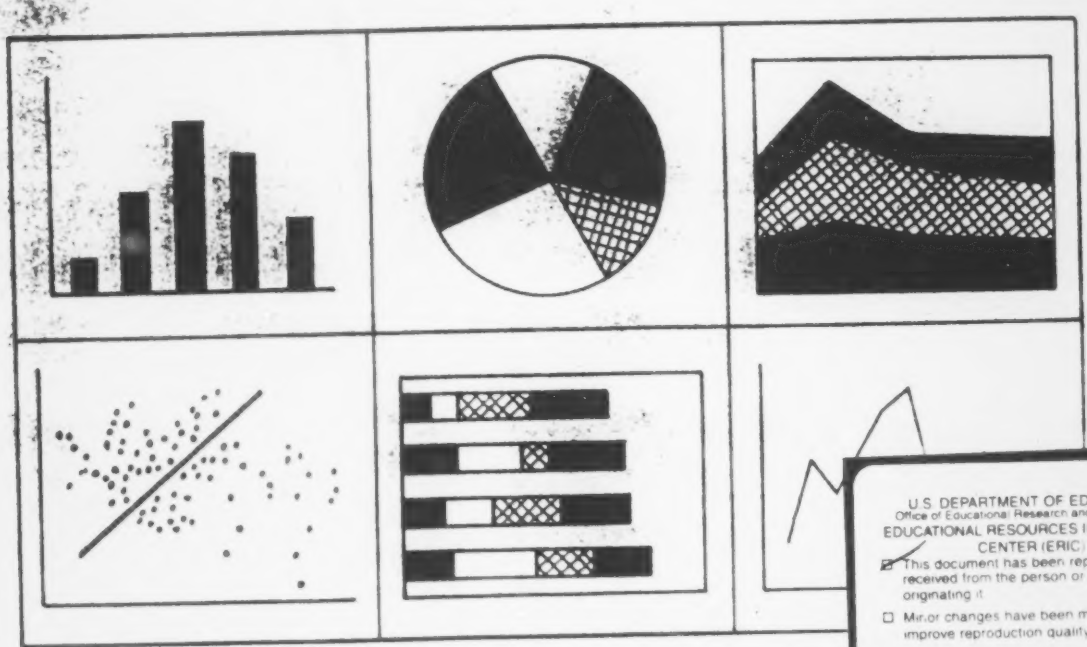
Working Paper Series

Formulating A Design for the ECLS:

A Review of Longitudinal Studies

Working Paper No. 97-24

August 1997



U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

TM028134

U.S. Department of Education
Office of Educational Research and Improvement

Formulating A Design for the ECLS:

A Review of Longitudinal Studies

Working Paper No. 97-24

August 1997

Contact: Jerry West
Data Development and Longitudinal Studies Group
(202) 219-1574
e-mail: jerry_west@ed.gov

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Ricky T. Takai
Acting Assistant Secretary

National Center for Education Statistics
Pascal D. Forgione, Jr.
Commissioner

Data Development and Longitudinal Studies Group
Mary J. Frase
Associate Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, DC 20208

Suggested Citation

U.S. Department of Education. National Center for Education Statistics. *Formulating a Design for the ECLS: A Review of Longitudinal Studies*, Working Paper No. 97-24, by Patricia J. Green, Lisa A. Hoogstra, Steven J. Ingels, Harrison N. Greene, and Patricia K. Marnell. Project Officer, Jerry West. Washington, D.C.: 1997.

August 1997

Foreword

Each year a large number of written documents are generated by NCES staff and individuals commissioned by NCES which provide preliminary analyses of survey results and address technical, methodological, and evaluation issues. Even though they are not formally published, these documents reflect a tremendous amount of unique expertise, knowledge, and experience.

The ***Working Paper Series*** was created in order to preserve the information contained in these documents and to promote the sharing of valuable work experience and knowledge. However, these documents were prepared under different formats and did not undergo vigorous NCES publication review and editing prior to their inclusion in the series. Consequently, we encourage users of the series to consult the individual authors for citations.

To receive information about submitting manuscripts or obtaining copies of the series, please contact Ruth R. Harris at (202) 219-1831 or U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 555 New Jersey Ave., N.W., Room 400, Washington, D.C. 20208-5654.

Samuel S. Peng
Acting Director
Statistical Standards and Services Group

Formulating a Design for the ECLS:

A Review of Longitudinal Studies

Prepared by:

Patricia J. Green
Lisa A. Hoogstra
Steven J. Ingels
Harrison N. Greene
Patricia K. Marnell

National Opinion Research Center (NORC)
at The University of Chicago

Prepared for:

U.S. Department of Education
Office of Educational Research and Development
National Center for Education Statistics

Prepared under contract RN940940001.
The views expressed are those of the authors;
no endorsement by the government should be inferred.

August 1997

Preface

The **Early Childhood Longitudinal Study (ECLS)** is a new study that will focus on children's early school experiences beginning with kindergarten. The ECLS is being developed under the sponsorship of the U.S. Department of Education, National Center for Education Statistics (NCES), with additional financial and technical support provided by the Administration of Children, Youth, and Families, the U.S. Department of Education's Office of Special Education Programs and Office of Indian Education, and the U.S. Department of Agriculture's Food and Consumer Service. Approximately 23,000 children throughout the country will be selected to participate as they enter kindergarten and will be followed as they move from kindergarten through 5th grade. Base-year data will be collected in the fall of 1998, with additional spring follow-up data collections scheduled for 1999 through 2004. Information about children's neighborhoods, families, schools, and classrooms will be collected from parents, teachers, and school administrators.

Because of the magnitude and complexity of the ECLS, NCES has set aside an extended period of time for planning, designing, and testing the instruments and procedures that will be used in the main study. NCES and its contractor, the National Opinion Research Center, are using this time to examine a variety of issues pertaining to the sampling and assessment of young children and their environments. The design phase of the study will culminate in a large-scale field test during the 1996-97 school year.

NCES has sought the participation and input of many individuals and organizations throughout the design phase of the ECLS. The participation of these individuals and organizations has resulted in a set of design papers that identify policy and research questions in early education, map the content of the ECLS study instruments to these questions, and explore and evaluate different methods for assessing the development of children and for capturing data about their homes, schools, and classrooms.

This paper is one of several that were prepared in support of ECLS design efforts. The information on the studies described in this paper were current at the time the paper was written. We recognize that work on some of the studies has moved forward since that time. It is our hope that the information found in this paper not only will provide background for the development of the ECLS, but will be useful to researchers developing studies of young children and their educational experiences.

Jerry West
ECLS Project Officer

Jeffrey A. Owings
Program Director
Data Development and Longitudinal
Studies Group

Table of Contents

Foreword iii

Preface vi

1.0 Implications of Prior Research for ECLS **1**
- 1.1 Cognitive Assessment 1
- 1.2 Social and Emotional Measures of Children's Development 8
- 1.3 Environments 13
- 1.4 Study Design and Sampling 16
- 1.5 Administration 18
- 1.6 Summary 20
- References 22

2.0 Study Summaries **25**
- 2.1 Beginning School Study 25
  - References 32
- 2.2 Children of the National Longitudinal Survey of Youth 34
  - References 50
- 2.3 Greensboro Early Schooling Study 52
  - References 61
- 2.4 Prospects: The Congressionally Mandated Study of Educational Growth and Opportunity 62
  - References 69
- 2.5 The D.C. Early Learning and Early Identification Longitudinal Study . . . 70
  - References 77
- 2.6 National Education Longitudinal Study, 1988 (NELS:88) 78
  - References 89
- 2.7 The Canadian National Longitudinal Survey of Children 91
  - References 100
- 2.8 The National Survey of Children 101
  - References 105
- 2.9 National Child Development Study 106
  - References 109

List of Tables

Table 1: Type of Cognitive Assessment Used 2
Table 2: Reported Correlations among Assessments of Entering Kindergarten Students: Greensboro Study 5
Table 3: Correlations among Assessments Used with Children of NLSY79 6
Table 4: Type of Social and Emotional Measures of Child Development Used 9
Table 5: Measures of Children's Environments 15
Table 6: Base year through fourth follow-up--NELS:88 components 79

8 BEST COPY AVAILABLE

1.0 Implications of Prior Research for ECLS

Nine studies are reviewed here, each of which provides some unique design features that ECLS may wish to emulate:

- Beginning School Study (BSS)
- Children of the National Longitudinal Survey of Youth (NLSY79)
- Greensboro Early Schooling Study (Greensboro)
- Prospects: The Congressionally Mandated Study of Educational Growth and Opportunity (Prospects)
- District of Columbia Early Literacy and Early Identification Study (DC)
- National Education Longitudinal Study of 1988 (NELS:88)
- The Canadian National Longitudinal Survey of Children (NLSC)
- The National Survey of Children (NSC)
- British National Child Development Study (NCDS)

The order in which these studies are reviewed reflects their relevance to ECLS. Five out of the first six studies listed are school-based studies. The exception is the Children of the National Longitudinal Survey of Youth, a home-based study using direct assessment that has produced a rich body of empirical data.

In reviewing the design and results of these studies, several cross-cutting questions emerged that will need to be addressed and decided. This chapter articulates several of these questions concerning the cognitive assessments, the social and emotional measurements, and the measures of environment. Issues concerning overall study design are discussed at the end of the chapter.

1.1 Cognitive Assessment

Table 1 presents information on the types of assessments that have been used in past surveys. The Peabody Individual Achievement Tests (PIAT) and Peabody Picture Vocabulary Tests (PPVT) have been used together in several studies: the National Longitudinal Study of Labor Market Experience—Youth Cohort (NLSY79), the Greensboro Early Schooling Study, and the British National Child Development Study (NCDS). The Canadian National Longitudinal Study of Children (NLSC) used the PPVT. Each of these surveys included preschool children in the sample. In addition, since NLSY79, NCDS, and NLSC were in-home studies, the use of individual assessments was a practical necessity. Table 1 also distinguishes between the version of the PIAT or PPVT used -- the original PIAT was published in 1970, and the revised PIAT, or PIAT-R, in 1989. The original PPVT was published in 1970, and the revised version, PPVT-R, in 1981; a new version, PPVT-3, is in development.

Table 1: Type of Cognitive Assessment Used

Study:	Vocabulary	Reading	Mathematics	Other	Academic Performance
Beginning School Study	California Achievement Test - Verbal (phonic analysis, vocabulary, comprehension, language, structural analysis)		California Achievement Test - Quantitative (computation and concepts)		Grades on Report Cards
Prospects	Comprehensive Test of Basic Skills	Comprehensive Test of Basic Skills	Comprehensive Test of Basic Skills		Teacher's Reports of Grades: "Mostly A's, Mostly B's, etc."
DC Longitudinal Study	Comprehensive Test of Basic Skills	Comprehensive Test of Basic Skills	Comprehensive Test of Basic Skills	CTBS Science, Social Studies	Grades from Progress Reports Competency Based Checklist
Greensboro Early Schooling Study	PPVT-R	PIAT-R Reading Recognition	PIAT-R Math	PIAT-R General Information	Grades from Progress Reports
NELS:88	None	NELS Reading Comprehension	NELS Mathematics	NELS Science NELS Social Science	Transcripts; Teacher Ratings
NLSY79 Child and British NCDS	PPVT-R Form L	PIAT Reading Recognition & Comprehension	PIAT Math	None	Parental Reports of School Performance
Canadian Longitudinal Survey of Children	PPVT	None	None	None	Parent and Teacher Reports of Performance
National Survey of Children	None	None	None	None	Parent, Teacher, and Child Reports of Performance

Several of the school-based studies used standardized achievement tests designed to be sensitive to the school curriculum. Prospects used the Comprehensive Test of Basic Skills, a battery of tests administered over a three-day period; the DC study also used CTBS with third-

grade students.[1] Conducted in Baltimore, the BSS used the California Achievement Tests that were administered by the school district beginning in first grade. Researchers at Educational Testing Service (ETS) and elsewhere worked together to design the criterion-referenced achievement tests used in NELS:88. Note that all of these achievement tests were administered to groups of children, and all were designed for use with children in first grade or higher, but group-administered exams will be impractical for five-year olds.

Several issues became apparent in reviewing the types of assessment instruments used in each of these studies and the research results based on the findings. These issues include:

- the advantages and disadvantages of curriculum-sensitive assessments and/or criterion-referenced assessments;
- the need for adaptive testing;
- the intercorrelations among measures;
- the importance of modeling growth during both the academic year and summer;
- the relationship between assessments and school performance.

Curriculum-sensitive assessments. ECLS seeks to measure children's achievement in school. Consequently, a major disadvantage of using instruments that are not tied directly to the curriculum is that such assessments are likely to be closer to ability tests than curriculum-sensitive achievement batteries. Such criticisms were lodged at High School and Beyond, the NCES longitudinal study of high school students launched in 1980. As a result, for NELS:88, teachers and curriculum specialists created tests with specifications for curriculum content.

Development of curriculum-sensitive tests may also make it easier to define criteria for mastery levels or proficiency levels that are midpoints on the expected growth curve. Experience with NELS:88 demonstrates that developing criterion-referenced markers within subject areas aids in interpretation of the data, because policy implications are clear when research can be expressed in terms of mastery or proficiency levels.

Assessments designed in this way also have significant disadvantages. To the extent that there is great diversity in curricula across schools and districts, developing a national curriculum-sensitive test is inadvisable because school-specific error is apt to be high. However, since the primary grades do emphasize the development of basic reading, writing, and mathematics skills, it may be possible to develop an assessment that reflects the elementary school curriculum. At the same time, since the curriculum in early grades focuses on basic skills, instruments that assess

[1] The DC study used the Metropolitan Reading Test to measure reading readiness among kindergarten students.

those general skills (such as the Peabody instruments) may be, in fact, curriculum sensitive. Additional empirical work will be necessary to examine the degree to which this is true.

A second disadvantage of the curriculum-sensitive achievement tests currently available is the time necessary for administration. The Comprehensive Test of Basic Skills is administered over the course of three days. Depending on grade level, the California Achievement Tests take between 1.5 and 5.5 hours, although short forms (taking approximately 2.5 hours) of the longer tests are available. NELS:88 included four subject area tests and was designed to be administered in 85 minutes; the math and reading test took approximately 50 minutes. Clearly, because the early waves of ECLS will be administered by interviewers in a one-on-one setting, the time that can be spent in conducting the assessment is short. While an abbreviated assessment cannot comprehensively represent the curriculum in the sense of supporting subscales, a representative selection of items from the curriculum can be used to measure achievement growth over time, and mark movement on a behaviorally-anchored skill and knowledge hierarchy.

Adaptive testing. To assess individual change over time in a longitudinal survey, it is necessary to use measures that have proven to be extremely precise and reliable. A meaningful level of precision can only be achieved when a respondent has the opportunity to complete several items that are within his or her range of knowledge. Unless the assessment instrument is very long, or the group being assessed is very homogenous, the potential for floor and ceiling effects is very real. From the studies reviewed here, we know that ECLS will certainly include a diverse group of kindergarten students. Morrison, Griffith, & Williamson (1993) report that the span of vocabulary evident in his sample of kindergarten students in Greensboro, North Carolina, ranged from that of a two-year-old to that of a nine-year-old.

One way of decreasing the administration time while still maintaining a sufficient level of precision is to turn to adaptive testing. The computer-assisted personal interviewing version of the PIATs and PPVTs developed for use in NLSY79 offer a useful model for ECLS. The Peabody instruments are designed as adaptive tests: the interviewer chooses the first item based on the child's age and then works backwards and forwards until a floor and a ceiling are established. An alternative model for ECLS would be NELS:88's use of test forms in the follow-up rounds that are tailored to a particular student's ability level based on prior results or a "duplex design" with a short routing test. Multilevel tests are often desirable to avoid floor and ceiling effects in longitudinal measurement. Using a single set of items for students with different abilities and achievement levels can seriously inflate the error of measurement.

Intercorrelations among measures. One of the goals set forth in the initial proposal for ECLS is that achievement in each area, for example in language skill or mathematics, be measured as distinctly as possible, drawing only on skills in the single area being assessed. Since most of the studies that are reviewed here include assessments of student achievement in mathematics and reading, one question to be addressed is the extent to which these measures of language and math skills are correlated. Published results from the Greensboro Study and NLSY79 are available.

Morrison and his colleagues (1993) report that, among entering kindergarten students, the PIAT and PPVT scores were fairly strongly correlated with each other, as well as with a measure of IQ used in the study. Table 2 presents correlations reported by the principal investigator. Analysis of the assessments administered to entering kindergarten students indicates that correlations range from .52 for receptive vocabulary and reading recognition scores to .78 for scores on the receptive vocabulary and general knowledge assessments. Morrison and his colleagues interpret these findings as indicative of general differences in individual skill levels (i.e., children who score relatively low in one domain tend to score relatively low across all domains).

However, correlations across domains may reflect the instruments' failure to assess skills in only one domain. Children with poor verbal comprehension skills would score lower on all assessments that rely on complex verbal instructions or have items that are embedded in verbal contexts.

Table 2: Reported Correlations Among Assessments of Entering Kindergarten Students: Greensboro Study

PIAT-R, PPVT-R, Test Source	Receptive Vocabulary	Reading Recognition	Cultural Knowledge	Mathematics
Receptive Vocabulary				
Reading Recognition	0.52			
Cultural Knowledge	0.78	0.57		
Mathematics	0.62	0.61	.65	

Source: Morrison, Griffith & Williamson (1993)

Published results of the NLSY79 data, presented in Table 3, also indicate that reading and math scores are correlated. Correlations between the PPVT (available for only 10-11 year olds in 1990) and math, reading recognition, and reading comprehension are modestly high: the PPVT has a .52 correlation with the PIAT math, a .55 correlation with PIAT reading recognition, and .59 correlation with reading comprehension. Data from 5-7 year olds in 1986 show a correlation between math and reading recognition of .46, and a correlation of .45 between math and reading comprehension. In Table 3, cells marked "NA" indicate "correlation with self, not applicable". Cells marked with a "U" mark correlations that are unavailable from the *NLSY Child Handbook*.

Table 3: Correlations among Assessments Used with Children of NLSY79

	PIAT - Math			PIAT - Reading Recognition			PIAT - Reading Comprehension			PPVT
	86	88	90	86	88	90	86	88	90	90
1986 PIAT - Math										
All ages	NA	.59	.57	.57	.50	.50	.55	.51	.45	U
5-7 years		.54	.52	.46	.43	.43	.45	.42	.42	
8-9 years		.61	.59	.59	.48	.48	.57	.50	.43	
10 and over		.65	.63	.63	.59	.59	.60	.59	.49	
1986 PIAT - Reading Recognition										U
All ages	.57	.53	.52	NA	.71	.70	.81	.63	.58	
5-7 years	.46	.43	.39		.57	.55	.82	.53	.52	
8-9 years	.59	.58	.61		.82	.77	.78	.70	.65	
10 and over	.63	.60	.60		.75	.85	.78	.64	.55	
1986 PIAT - Reading Comprehension										U
All ages	.55	.49	.50	.81	.62	.62	NA	.60	.52	
5-7 years	.45	.40	.37	.82	.56	.55		.52	.47	
8-9 years	.57	.55	.58	.78	.65	.61		.65	.54	
10 and over	.60	.53	.56	.78	.59	.72		.55	.52	
1986 PPVT	U			U			U			
All ages		.44	.41		.40	.44		.42	.44	
3-5 years		.40	.37		.29	.38		.31	.40	
6-7 years		.43	.39		.41	.46		.45	.47	.66
8-9 years		.48	.51		.50	.45		.51	.51	
10 and over		.59	.61		.62	.76		.62	.61	
1990 PPVT 10-11 year olds			.52			.55			.59	

Source: Baker, Keck, Mott, & Quinlan, *NLSY Child Handbook, Revised Edition: A Guide to the 1986-1990 National Longitudinal Survey of Youth Child Data*. Center for Human Resource Research, Ohio State Univers ity. Pages 301 and 325.

These within-year correlations across tests are similar in magnitude to the cross-year correlations on single tests. Reading comprehension among 5-7 year olds in 1986 has a .52 correlation with reading comprehension among the same group two years later in 1988. The mathematics score among 5-7 year olds in 1986 has a .54 correlation with the 1988 mathematics score. In comparison, the 1986 correlation between math and reading comprehension is .45.

The observed pattern of results could reflect children's general intelligence. As intelligence theorists have suggested in the past, children's abilities across domains may be very similar. However, the pattern of higher correlations among older children than younger children does not seem to support this explanation: the correlations across areas and across years in NLSY79 are lower for the 5-7 year old age group than for older children. That is, the correlation between math and reading comprehension is higher for those 13 and over than for the 5-7 year old age group.[2] If the correlation across tests was a function of "generalized intelligence," one would expect that it would be stable across time. The alternative explanation is that this pattern is a function of the assessment itself. One plausible interpretation for these findings is that all of the assessments (math, general knowledge, reading, and vocabulary) use verbal skills and that the source of the correlation is verbal ability. A definitive answer to this question awaits further empirical investigation.

Academic year and non-academic year growth. Several studies are finding that the rate of cognitive growth during the school year is similar among more- and less-advantaged school children. However, the rate of cognitive growth experienced by children from the two groups during the summer is markedly different. Data from Prospects (Rock, 1994) and the BSS (Entwisle and Alexander, 1992) indicate that children from more advantaged backgrounds continue to make cognitive gains during the summer, while the rate of growth among children from less-advantaged backgrounds comes to a halt during the summer months.

The implications of this for ECLS are clear. If assessments are conducted once each year after kindergarten, the impact of schools on less-advantaged students will be consistently and systematically underestimated. The slope of the line plotting cognitive gains estimated for more advantaged students will be steeper than the slope estimated for less-advantaged students, even though, based on the studies reviewed here, we know that the lines plotting the true academic-year gains are parallel. Thus, researchers and policymakers are apt to conclude, based on ECLS data, that schools do a better job with richer students than with poorer students—a plausible but probably incorrect inference with enormous policy implications.

School performance. Student performance in school is nearly always evaluated. However, especially in kindergarten, there is wide variation across schools in the dimensions of student performance that are evaluated, and no single metric is typically used for these measurements. Several of the studies reviewed used measures of school performance as outcome measures in analysis: the BSS used report card grades and the DC study used progress reports. Other studies, such as Prospects, asked children or informants to report on their grades (i.e., whether they were "mostly A's, mostly B's", and so on). Information about retention and promotion was gathered in almost all studies reviewed.

[2] Alexander and Entwisle (1988) also find that fall CAT scores are more predictive of spring CAT scores for older than younger students (second grade students compared to first-grade students).

Using data from report cards, progress reports, or transcripts will be difficult in ECLS. Studies that have used this type of data in elementary school have been limited to single school systems with common grading practices. Because ECLS has a national scope, a suitable coding system will have to be developed to standardize measures. Because of the central importance of school performance for ECLS, careful pilot development work must be conducted to determine the type of school performance data that can be reliably collected for primary school students.

Information on children's attendance, grade-retentions, suspensions or expulsions, and referrals for behavioral or learning handicaps have been obtained in past school-based studies from school records, while most of the household studies relied on parental reports. During the ECLS field test, we will be able to examine the types of school records available, and the ease of coding these records across schools.

Some constituent groups interested in ECLS will advocate using systematic performance assessments. Such assessments have been used in past research with smaller samples, although none of the large-scale surveys reviewed here included performance assessments. However, the DC study did include a "competence assessment," that gave teachers a self-administered questionnaire for each child in the study that asked whether or not the student was able to complete the task named (for example, "multiply whole numbers"). The assessments were designed to be sensitive to the district's curriculum, and to measure what children had learned in school. The feasibility and utility of developing such assessment checklists could be explored for ECLS.

1.2 Social and Emotional Measures of Children's Development

Table 4 presents information on the types of social and emotional measures used in the reviewed studies. For ease of presentation, the measures were classified into three primary groups: children's adaptive behavior and social skills, children's self-concept, and children's expectations about their own performance.

Adaptive behavior and social skills. Children's ability to adapt to the school environment is generally considered to be an important factor influencing success in school. All of the studies reviewed examined some aspects of children's behavior. However, the constructs used varied across studies. Because adaptive behavior is often defined, in part, by social skills, it is necessary to discuss these two dimensions together.

Table 4: Type of Social and Emotional Measures of Child Development Used

Study:	**Adaptive Behavior/ Social Skills**	**Self-concept**	**Expectations**	**Other**
Beginning School Study	Personal Maturity Scale (Subset of Behavior Problems Index) (Teacher)	Dickstein (23 item) Scale: (Character, Responsibility, Academic, Athletic, and Appearance)	Expected Grades	
Prospects				
DC Longitudinal Study	Vineland Adaptive Behavior Scales (Teacher)			
Greensboro Early Schooling Study	Cooper-Farran Behavior Rating Scale (Teacher)	Pictorial Scale of Perceived Competence and Social Acceptance for Young Children		
NELS:88		Items measuring locus of control; self-esteem; and, for 1990, Marsh's Academic Self-Concept Scale		
NLSY79 Child and British NCDS	Behavior Problems Index (Parent) Interviewer observations	Self-Perception Profile for Children		How My Child Usually Acts (Temperament)
Canadian Longitudinal Survey of Children	Vineland Adaptive Behavior Scales (Parent)	Child Questionnaire (10-11 year olds): Items from the Marsh Self-Description Questionnaire		Infant Characteristics Questionnaire (Temperament)
National Survey of Children	Behavior Problems Index (Parent and Teacher)	Items in the Child Interview (child's feelings about self)		

NLSY79, NCDS and the BSS adapted items originally used in the NSC. The 12-item Personal Maturity Scale used in the BSS contains a subset of items from the Behavior Problems Index, but for balance also uses four items about positive behaviors, which are said to form a single factor (Alexander and Entwisle, 1988). Data from the 28-item Behavior Problems Index are analyzed as six separate subscales: antisocial, anxious/depressed, headstrong, hyperactive, immature dependency, and peer conflict/social withdrawal. Analyses of data from both studies demonstrate a negative association between these behavior scales and tested achievement (Baker,Keck, Mott & Quinlan 1993; Alexander and Entwisle, 1988). Findings from the DC Early Learning and Early Identification Study also demonstrate a negative correlation between "maladaptive behavior," as measured by the Vineland Adaptive Behavior Scales, and tested achievement (Marcon, 1994).

One of the major issues raised in reviewing all studies is the relative importance of measuring positive or negative behaviors. Data clearly show the importance of including some measure of behavior problems or maladaptive behavior. The studies reviewed here provide less guidance concerning the usefulness of measuring positive behavior. The exception to this is the DC study which used the Vineland Adaptive Behavior Scales, a set of subscales that measure communication skills, daily living skills, social development, and motor development, as well as maladaptive behavior. Findings from that study suggest that there may be a relationship between positive dimensions of adaptive behavior, as measured in kindergarten, and the likelihood of later retention (Marcon, 1994: 100).

In a review of the research on the measurement of social competence, adaptive behavior, and learning dispositions, Meisels, Atkins-Burnett, and Nicholson (1995) argue for the importance of measuring positive social behaviors, but advise against the use of existing adaptive behavior scales, including the Vineland. They note that social skills such as perspective or role taking, social judgment, and social problem-solving are not assessed on the adaptive behavior scales. Because these skills are a crucial factor in contributing to differences in achievement, they argue that ECLS should use a rating scale that measures these skills. Based on their review of the literature, they recommend the use of an adapted version of the Social Skills Rating System (SSRS), an instrument that primarily samples the area of social competence but also has some overlap with the adaptive behavior scales. They note that while the SSRS does include items that measure positive functioning, it does not address these areas of functioning in detail. We may thus want to consider adding items in this area.

Consideration of the relationship between behavior and achievement prompts discussion of a related issue, school readiness. The association between behavior problems and achievement may stem from intrapersonal factors, such as an inability to concentrate, or from a poor fit between the person and environment. Teachers undoubtedly have varying standards for acceptable classroom behavior; a student who has difficulty in one classroom may not experience the same problems in another classroom. Thus, it may be important to measure social competence and adaptive behavior in several environments, both at home and at school.

Child self-concept. Children's perceptions of their own competence can play an important role in their success in school. For many children, the first exposure to critical evaluation and comparative judgments regarding their performance comes when they enter school. How school experience alters children's valuations of their own competence is a question of considerable importance. While parents, teachers, and peers undoubtedly influence children's perceptions of their own competence, the ways in which children respond to others' evaluations, and the extent to which they value or discount their judgments, can serve to moderate the effects of such influences.

Most of the studies reviewed here included some measure of children's self-concept, but as shown in Table 4, no two studies used the same instrument. The NLSY79 child survey used two subscales of Harter's Self-Perception Profile for Children to measure children's perceived competence in the academic skills domain and their sense of global self-worth. (The scales are used only with children who are 8 years or older.) The Greensboro study used an instrument designed for younger children, Harter and Pike's Pictorial Scale of Perceived Competence and Social Acceptance for Young Children. Subscales of that instrument used in the study included cognitive competence, physical competence, peer acceptance, and maternal acceptance. Although the reliabilities for the overall scale are reasonably high, the reliabilities for the subscales are modest (Harter and Pike, 1984), and reliabilities for differing age groups are not similar. In addition, the competence subscales (cognitive and physical) have been found to be less reliable than the social acceptance subscales (peer and maternal). Results from Frazier and Morrison's (1994) extended-year study raise some concern about the reliability of these scales: self-reported levels of cognitive competence decreased over the summer, even among students who were in extended-year programs.

The BSS used a set of 23 items developed by Dickstein (1972), which were designed to tap self-concept in five areas: character, responsibility, academic competence, athletic competence, and appearance. Analysis of data from these items suggests that the dimensions become more clearly differentiated over time; the correlations across dimensions are higher among first-grade children than second- or fourth-grade children (Pallas, Entwisle, Alexander, and Weinstein, 1990). Some differences in self-esteem were evident by gender and race; however, social class differences in esteem were negligible.

Measuring the self-concept of kindergarten-age children is problematic. A primary problem is that self-concept scales rely on self-report. As Meisels and colleagues note, "problems with reliability in self-report instruments with young children are legion" (1995, p. 14). Most of the measures are reliable only when used with older children. This may indicate that more work is needed to develop a reliable measure of self-concept that can be used with young children. However, it is also possible that the notion of self-concept has limited applicability to children in this age group. A second problem is that self-concept scales measure only one dimension of socio-emotional development and would need to be used in conjunction with other measures. Based on these considerations, Meisels and colleagues (1995) recommend that ECLS use a more comprehensive set of measures to assess social competence and emotional well-being.

Performance expectations. The BSS is unique in measuring performance expectations: parental and student expectations are key variables in the model set forth in that study. Both parents and students were asked how well they thought the student would do in school by querying them about the grades expected in reading and mathematics. Alexander and Entwisle's (1988) research raises several important questions about how expectations should be measured and modeled. Research based on these data highlight the importance of viewing schooling as a dynamic process, where what is learned and how one is evaluated affect subsequent expectations, learning, and evaluation.

The way that expectations are used in the study underline the importance of performance expectations. Expectations are seen as key mediating variables affecting school outcomes. Students' expectations are shaped by their abilities, parental expectations, and past performance. Analysis using BSS data demonstrates that grades from report cards affect students' subsequent performance expectations.

A related issue that was not fully addressed in the materials reviewed is how feedback on performance affects students' self-appraisals and behavior. Literature in education has long noted the recursive nature of both success and failure, but only limited empirical research has been conducted that examines the mechanisms by which one failure fuels another or one success leads to the next. The studies conducted in DC and Baltimore were both designed, in part, to examine this process, paying particular attention to the outcomes associated with grade retention.

The BSS demonstrates that data on students' expectations can be obtained from children as young as six years of age. However, the procedures used by BSS to obtain these data appear to have been fairly time consuming. Many beginning first graders were unfamiliar with how they would be evaluated and interviewers had to explain how grades were assigned and what they meant before they could ask the children what grades they expected to get on their next report card. Because ECLS will be operating under tight time constraints, and will be dealing with schools that have a variety of grading practices, the procedures used for BSS have limited usefulness for ECLS.

Self-reports by children. Many of the studies reviewed here included a child survey component, either a self-administered questionnaire or an interview, that asked questions of the child directly. BSS included a personal interview beginning in first grade that includes questions about their expectations about grades in mathematics and reading (Entwisle and Hayduk, 1982). The Greensboro study did not include survey items, but did ask kindergarten children to complete Harter and Pike's Pictorial Scale of Perceived Competence and Social Acceptance for Young Children. Children ages 7 to 11 were interviewed in the first wave of the NSC.

Prospects asked third-grade children to complete a self-administered questionnaire but did not ask first-grade children to do so. In Prospects, the third grade questions were read aloud in the classroom and two assistants "floated" through the room helping children who needed assistance. Other studies relied on children's self reports only for those age ten and older. The NLSY79 does not give the Child Self-Administered Supplement to children under ten; and

Statistics Canada does not ask children under ten to complete the self-administered questionnaire that is part of the NLSC.

In ECLS, only a limited amount of time is available for direct contact with children and most of this must be dedicated to a direct assessment of cognitive skills. No plans have been made for conducting interviews with kindergarten respondents. However, prior research does suggest that a few questions might be useful, especially in areas such as self-concept. While the reliability of young children's responses are suspect, there are certain areas, such as performance expectations, in which children's answers are probably more reliable than responses of proxy informants.

Consistency of parent and teacher reports. ECLS will need to rely on proxy respondents to gather information on children's behavior and social skills. This has been done previously; all of the studies reviewed relied on reports from either parents or teachers to assess children's adaptive behavior and social skills. One study, the NSC, asked both teachers and parents similar questions about children's behavior. In that survey, the parent and child interviews and teacher questionnaires were carefully articulated, making it possible to identify convergent or divergent views on children's relationships to family members, peers, and teachers, as well as other measures of behavior and academic performance.

The use of multiple proxy informants reinforces the notion that questions measuring children's behavior and social skills should be designed to be context-sensitive and should not attempt to measure "global" aspects of children's behavior or skills. This may be especially important in assessing children's behavior. One of the opportunities that ECLS will provide is to ask parents and teachers about children's behavior at home and at school, and about their own standards concerning appropriate behavior.

1.3 Environments

The studies reviewed here are all based on survey research methodology. As such, they provide a solid source of survey items that measure demographic characteristics and attitudes of respondents. However, in comparison to observational studies, these surveys include only limited measures of context and interaction. Thus, the synopsis presented here is limited to methodological issues in measuring environments that cut across the reviewed studies. Table 5 presents the sources of contextual data gathered in each of the studies reviewed. A more exhaustive discussion of the dimensions of home and school environment will be included in the content outlines for the teacher, parent, and school questionnaires (see Dauber et al., An Outline of Contextual Measures for the Early Childhood Longitudinal Study).

Observation vs. respondent reports of environments. The design of ECLS calls for the collection of data about the environments in which children develop—school, home, and neighborhood. Because of the large number of respondents included in national surveys, it is typically impossible to incorporate systematic repeated observations of classroom or home environments. However, as an alternative to relying solely on the reports of parents and teachers,

some studies have relied on observations recorded by the interviewer at the time of data collection. For example, NLSY79 asks interviewers to complete some items from the short-form of the HOME inventory; mothers are asked to complete other items. Unfortunately, no data are available on the inter-rater reliability of the items completed by the interviewers on the short-form. While cross-year correlations are quite high (there is a .54 correlation between the 1986 and 1988 composite HOME scores), these composites are based on maternal and interviewer reports, and every attempt was made to use the same interviewers in both rounds of the study. The ECLS field study might profitably explore the quality of single time-point interviewer observations.

The impact of school environments. One of the major policy issues ECLS will address is how schools promote or hinder student learning in the early grades. Although ECLS will not attempt to evaluate particular programs, it will examine the impact of school environments on student learning. Results of the study may be helpful in identifying ways in which schools, classrooms, and instructional practices might be modified to improve student academic achievement.

To the extent that ECLS examines the appropriateness of classroom environments or school policies in promoting student achievement, it will be necessary to develop direct measures of the hypothesized impact of such environments or policies. Rather than inferring information from student outcomes, conclusions regarding the developmental appropriateness of particular programs should be based on measures that are developed specifically for that purpose. An illustration of the importance of this is in order. Researchers analyzing data from the DC study of early learning programs found that children who attended a prekindergarten program with a strong academic focus were more likely to earn lower grades and exhibit behavior problems in fourth grade than their peers who attended programs oriented toward socio-emotional development (Marcon, 1994: ix-x). This finding is extremely important and would be far more convincing if the data could be used to shed light on why this is so; for example, did the academic focus produce early failures that initiated a pattern of low expectations and low achievement?

It is necessary to specify in advance of the study the *process* by which these environments affect students. The BSS is a strong model in this regard, because the investigators clearly articulated a model of school achievement and developed measures of each stage of the process: entry knowledge, parental expectations, student expectations, student grades, and tested achievement. A dynamic model of student learning is needed, one that incorporates measures of the interplay between the individual and the environment.

Table 5: Measures of Children's Environments

Study:	School	Home	Neighborhood	Other
Beginning School Study	Teacher Questionnaire	Parent Questionnaire		
Prospects	Teacher Questionnaire	Parent Questionnaire		
DC Longitudinal Study	Teacher Questionnaire			
Greensboro Early Schooling Study	Early Childhood Environments Rating Scales (ECERS) [CISSAR to be used in future]	Parent Questionnaire on: Family Literacy Knowledge and Beliefs Rules and Limits Family Organization Affective Climate		
NELS:88	Teacher Questionnaire	Parent Questionnaire		Census Information
NLSY79 Child and British NCDS		Parent Questionnaire HOME-Short Form Interviewer Observations		
Canadian Longitudinal Survey of Children		Parent Interviews	Parent Interviews: Sections from the Smicha-Fagan Neighborhood Questionnaire; Interviewer Observations: Items from the Neighborhood Cluster Observation Schedule	Parent Interviews: Social Provisions Scale Short Version (a measure of perceived social support)
National Survey of Children		Parent and Child Interviews; Interviewer Observations	Parent and Child Interviews	

Opportunity to learn and amount of instruction. Perhaps the single largest gap in the studies reviewed here is that none attempted to quantify the amount of instruction received by students. During the past several years, one of the major advances in educational research has been the focus on the opportunities for learning provided by schools and the amount of instruction received by each student. Children do not all have equal opportunities to learn. Research has demonstrated that course offerings vary across schools, that differences exist in the amount of material covered in courses in differing "tracks," and that differences exist in the amount of instruction provided to students within the same classroom but in different ability groups (Gamoran, 1986; Dreeben and Barr, 1988).

In generating instruments for ECLS, careful consideration should be given to developing reliable measures of the amount of instruction received by individual children. Similarly, some effort might be spent on developing parallel measures that could quantify the amount of parent-child interaction at home. Reliance on survey methods will make this task difficult, and it may not be possible. However, if successful, such a pioneering effort would make a real contribution to the field and should be explored further using nonsurvey studies for guidance.

Institutional effects. The importance of collecting student record data to supplement data from direct assessments was discussed earlier. It may also be useful to collect information on the types of student record information that are passed from one teacher to the next, because teacher expectations have long been known to influence student performance. Recent research exploring the ways in which ability groups influence student achievement hypothesized that student placements in previous years influenced how teachers made decisions regarding appropriate placement. Students' grades and performance in previous grades affect not only the student's expectations, but the teacher's expectation as well. An important variable to include in the ECLS model of school achievement may well be the extent to which teachers know students' school histories.

1.4 Study Design and Sampling

Inclusion of special populations. Little is known about the relationship between test validity and use of special accommodations for testing the handicapped, and there are many views on the desirability of full inclusion (Ysseldyke & Thurlow, 1993; Thurlow, Ysseldyke & Silverstein, 1993). Nevertheless, NELS:88 provides several observations: (1) eligibility can change over time--an important consideration in a longitudinal study that expects to freshen follow-up samples to make them grade-level representative; (2) there is much evidence that test inclusion and exclusion decisions on the part of school personnel may lack reliability or validity; (3) there clearly are means to obtain indirect information about individuals who cannot be directly assessed—information that may give evidence of important educational outcomes, and that, at the very least, provides a basis for estimating sample undercoverage biases and their impact on survey data.

Although several studies (NLSY79, Prospects, BSS, Greensboro, and DC) collected data on children's health conditions, limitations, and special referrals or placements, none of materials

that were available from these studies mentioned either the exclusion of children with disabilities from interviews or assessments or the use of special accommodations for children with identified disabilities. In the Interim Report for Prospects, Puma, Jones, Rock & Fernendez (1993) note that no exclusions of disabled or limited-English-proficient students were permitted. They also observe, however, that nonrespondents in the Prospects samples are more likely to be disabled or to have limited proficiency in English.

Language barriers. The experience of NELS:88, Prospects, and National Assessment of Educational Progress (NAEP) in including students with language barriers is informative for ECLS. NELS:88 was able to assess about half of students who had limited English proficiency (LEP). Overall, about 1.5 percent of the potential eighth grade sample had to be excluded for language reasons. However, the follow-back study of excluded students showed that of those who were excluded for language reasons, the majority were capable of completing survey forms two to four years later. This fact underlines the need to retain LEP/NEP (no English proficiency) students in longitudinal samples, even if they are unable to complete baseline tests. Of course, the number of NEP/LEP students is increasing, and is highest at the lower grades.

The 1992 NAEP identified 4 percent of the potential fourth grade sample as LEP and on this basis excluded from assessment 3 percent of the sample. At eighth grade, 3 percent were identified as LEP, and two thirds of them (2 percent of the sample) excluded. At twelfth grade, the 1992 NAEP identified 2 percent as LEP and excluded 1 percent of the sample for language proficiency reasons (Mullis, Dossey, Owen & Phillips, 1993). Current Population Survey data for 1989 (*Condition of Education 1992*) show that, of children 8 to 15 years old who are enrolled in school, 11.5 percent are language minority students (speak a language other than English at home) and 3.2 percent are LEP (by family self-report). Using a different reporting source (state education agencies), the 1993 Office of Bilingual Education and Minority Language Affairs (OBEMLA) LEP study (Henderson, Abbott & Strang, 1993) suggests that 5.6 percent of students nation-wide are LEP (but 19 percent for California and New Mexico). Again, LEP proportions are always somewhat higher in the lower grades and proportions are growing over time. For a kindergarten study in 1998-1999, the NELS:88 strategy of allowing NEP and some LEP students to be excluded is not likely to be acceptable.

For NLSY79, Spanish translations of several child assessment instruments were made available to respondents with limited proficiency in English. In 1986, a total of 354 children, age eight months or older, were assigned to bilingual interviewers. Of these cases, slightly more than 100 children were actually assessed in Spanish. By 1990, 52 children were assigned to bilingual interviewers. Of this number, 17 were actually assessed in Spanish. (See Baker et al., 1993, pp. 17-18 for the complete list of NLSY79 child assessment instruments for which Spanish translations are available.) In both NELS and Prospects, Spanish translations of questionnaires were made available to Spanish-speaking students and parents. Translations were not made available to respondents who spoke other languages. Puma et al. note that "Especially for the Parent Questionnaire, the Prospects design includes relatively high proportions of cases for which suitable instruments (e.g., versions of the questionnaires in several Asian languages) are not currently available" (1993, p. 12).

Smaller studies, such as BSS and Greensboro, did not include significant numbers of language minority students. Alexander and Entwisle (1988) note that, in 1980, less than one percent of Baltimore's population was Asian or Hispanic. Only seven Asian or Indian children were included in the original sample for BSS. Although the Greensboro followup studies, to be conducted in Greensboro, North Carolina, and Evanston, Illinois, may include a group of Hispanic students, the sample for the original study did not include any language minority students.

Since phone interviews with parents are currently planned for ECLS, the problem of translating protocols into other languages may be partially resolved through the use of bilingual interviewers. Given the range of languages encountered in NELS and Prospects, the task of locating and hiring interviewers who are proficient in these languages poses a significant challenge for ECLS.

1.5 Administration

Securing cooperation and consent. Of the studies reviewed, NELS:88 appears to provide the best model for securing cooperation from schools, teachers, and parents. Similar procedures appear to have been used in Prospects, but a description of these procedures was not available for review. The remaining school-based studies (the BSS, the Greensboro study, and the DC study) were limited to a single city or school district and are less useful as models for ECLS.

For NELS, several levels of cooperation were sought prior to soliciting a commitment to participate in the study from administrators of sampled schools. Endorsements were sought from key educational associations such as the Council of Chief State School Officers (CCSSO), the National Catholic Education Association (NCEA), and the National Association of Independent School (NAIS). Approval was then sought at the state and district levels for public schools and at comparable administrative levels for Catholic and other private schools. Principals or school administrators were approached only when approval for the study had been obtained at these higher levels.

Within each cooperating school, principals were asked to designate a school coordinator to serve as a liaison between NORC staff, who were conducting the study, and selected respondents (the school administrator, students, teachers, and parents). The school coordinator handled all requests for data and materials as well as the logistical arrangements for data collection on the school premises. School coordinators were also asked to help identify students whose physical or learning disabilities or limited proficiency in English would preclude their participation in the study. Coordinators were also responsible for the distribution of parental permission forms to sampled students. (Details of the procedures used in NELS base year and first followup are provided in Ingels, Scott, Rock, Pollack & Rasinski, 1994, Chapter 4).

Although NORC was extremely successful in securing cooperation at all levels for both the NELS base year and followup studies, it is clear that some of the procedures used in NELS will need to be modified for ECLS. Given the young age of the children who will participate in ECLS, more direct involvement on the part of field staff may be needed in obtaining parental consent for

children's participation. As noted above in the discussion on special populations, the inclusion of children with special needs will necessitate a close cooperation between field staff and school personnel in the identification and widest possible participation of children with disabilities or language limitations. Differences in mode of administration—individual versus group-administered assessments, telephone interviews with parents versus self-administered questionnaires, and the possible use of classroom observations—will also require an adjustment in administrative procedures.

Respondent burden and interviewer training. Two issues emerge with respect to respondent burden in the studies that have been reviewed. The majority of studies that focus on young children (age 8 or younger) emphasize the need to minimize the amount of time required for children's participation. NLSY79 limited direct assessments of children to approximately 30 minutes. In the Greensboro study, individual assessments of kindergartners were limited to two 20-30 minute sessions. These studies also emphasize the need for specialized training for child interviewers. For NLSY79, child interviewers participated in a two and one-half day training session that was geared toward developing child interviewing skills. They were also required to tape and submit their first actual child interviews to the NORC central office for case review. (See Baker et al., 1993, pp. 17-18 for a detailed description of interviewer selection and training.) Interviewers also completed an interviewer evaluation of testing conditions to gauge the child's attitudes towards testing and to record any events that may have interfered with the assessment.

Compensation for certain classes of respondents was used in at least two of the studies reviewed. In the proposed followup to the Greensboro study (Morrison, 1994), it was noted that teachers would be relieved of their classroom duties by a paid substitute so that they could fill out a behavior rating scale for each child in the study. NLSY79 directly compensated respondents for their participation. Each Youth respondent was paid ten dollars on completion of the main interview; NLSY79 mothers who participated in the child assessments were paid five dollars for each child assessed (Baker et al., 1993). For ECLS, teachers will be asked to complete rating scales for individual children as well a teacher questionnaire. Since teachers will assume a disproportionate share of the respondent burden in ECLS, some form of compensation may be warranted to encourage higher levels of participation by teachers.

Mobility and tracking. Since ECLS will follow both Head Start participants and kindergarten students as they progress through the early grades, student mobility and tracking is an issue of considerable importance to the study. Although some children will be enrolled in elementary schools that offer Head Start and kindergarten programs, and will remain at those schools in later grades, a sizeable number of children will change schools after the completion of the kindergarten or Head Start year. Moreover, as Bryant notes in a working paper on mobility for Prospects, "Parents of children in the lower grades move more often than parents in the higher grades. The parents on average are younger, not so likely to be settled in their jobs, and are not so strongly influenced by the children's needs and desires to be kept in the same school" (1991, p. 5).

None of the smaller school-based studies that were reviewed (BSS, Greensboro, DC) attempted to track students who transferred during the course of the study. NELS and Prospects thus offer the best models for student tracking. Since most students changed schools between the base year (8th grade) and first followup (10th grade) in NELS, NELS was confronted with the task of tracking the majority of students who participated in the study and offers a comprehensive set of procedures for tracking students. A detailed description of NELS tracking procedures can be found in section 4.6 of the *NELS:88 Base Year to First Follow-up Final Technical Report* (Ingels et al., 1994). After 18 weeks of tracking, NELS succeeded in locating 99 percent of the base year sample.

Since data collection will occur on a yearly basis for ECLS (after the initial year of the study), student tracking will need to be initiated earlier for ECLS than it was for NELS. Field staff will need to identify the schools children will attend the following year and to secure the permissions from these schools. There may also be considerable within-year mobility among children in kindergarten and Head Start programs. Consequently fall to spring tracking will also need to occur during the base year of the study. Because mobility is expected to be particularly high among Head Start participants, a tracking sample is currently planned for the Head Start pilot test.

1.6 Summary

Cognitive assessment. Several issues pertaining to the use of cognitive assessments emerged from the studies reviewed, and will be addressed in the design of ECLS. Specifically, ECLS will: (1) use curriculum-sensitive measures of achievement; (2) construct multilevel or adaptive assessments to shorten administration time and increase the precision of the assessments; (3) design assessments so that domains of learning are assessed as distinctly as possible and include assessments of children's verbal and nonverbal competencies; (4) evaluate the feasibility of conducting assessments in the fall and spring to measure, more precisely, children's learning in school; (5) explore several measures of school performance including the coding of report cards and the use of teacher ratings of student competencies.

Social and emotional measures. Two types of behavior were typically assessed in the studies reviewed and found to be related to student achievement: problem behaviors and adaptive behaviors. Ideally prosocial behaviors and social competence skills should also be assessed. While adaptive behavior scales measure some of these skills and behaviors, Meisels and colleagues (1995) note that social skills such as perspective taking, social judgment, and social problem solving are not assessed on adaptive behavior scales. One instrument, the Social Skills Rating System (SSRS), measures social competence skills and also has some overlap with the adaptive behavior scales. Because these skills contribute significantly to differences in achievement, we believe that ECLS should use a rating scale that measures these skills. Since parents and teachers may have different standards for appropriate behavior, it is important that ECLS use measures of social competence that are sensitive to differences in home and school environments, and can be asked of both parents and teachers.

Although the majority of studies reviewed included measures of the child's self-concept, results from these studies suggest that most measures of self-concept are reliable only with older children. Measures of self-concept may be introduced in the later years of the study, as part of a student questionnaire or interview. At that time, questions about students' performance expectations and adjustment to school may also be added.

Environments. ECLS will go beyond the surveys reviewed here in designing measures that can be used to characterize a child's learning environment at school. None of the large-scale surveys reviewed attempted to quantify the amount of instruction received by students. Research has demonstrated that course offerings vary across schools, that differences exist in the amount of materials covered in courses in differing "tracks," and that differences exist in the amount of instruction provided to students within the same classroom but in different ability groups. In generating instruments for ECLS, we will attempt to construct measures of the amount and content of instruction received by individual children, as well as measures of teachers' expectations for their students' performance.

Some of the studies reviewed used interviewer observations to supplement reports from respondents about children's home environments. While data are lacking on the reliability of interviewer-completed items, the quality of single time-point interviewer observations might be profitably explored in the ECLS field test.

Study design and sampling. The ECLS sample will be designed to represent all kindergarten students, regardless of their current placement or ability to participate in the initial round of the study. While ECLS staff may not be able to accommodate the needs of all students when designing the cognitive assessments, the study will collect information about all students from parents and teachers. Also, based on the experience of NELS:88 and other studies, students unable to participate in the initial round of ECLS because of language or other barriers may be able to participate in later rounds. Thus, no student will be excluded from the sample based on LEP status. Nor will individuals with disabilities be excluded from the sample, even if those disabilities preclude their participation in the direct cognitive asssessment.

Administration. ECLS will build on the procedures developed as part of NELS:88 to guide data collection activities. Project staff will also need to modify procedures in several areas, because of requirements unique to ECLS. First, respondent burden may inhibit participation, especially for teachers with large numbers of students in the sample. Thus, ECLS will examine ways to compensate teachers for the time they will spend reporting information about children in their classrooms. Second, analyses of pilot and field test data will be aimed specifically at estimating the amount of time children can spend engaged in the direct assessment. Finally, the pilot and field test will be used to provide information about student mobility and tracking in the early elementary grades. Because mobility is expected to be particularly high with the ECLS Head Start sample, a tracking study is currently planned for the Head Start pilot test.

BEST COPY AVAILABLE

References

Alexander, K.L. & Entwisle, D.R. (1988). Achievement in the first two years of school: Patterns and processes. *Monographs of the Society for Research in Child Development, 53*, Serial No. 218.

Baker, P.C., Keck, C.K., Mott, F.L., & Quinlan, S.V. (1993). *NLSY child handbook: A guide to the National Longitudinal Survey of Youth 1986-1990 child data.* Columbus, OH: Center for Human Resource Research, The Ohio State University.

Bryant, E.C. (1991). Mobility and dropouts. A working paper in partial fulfillment of the requirements for task 9 of Prospects: The Congressionally mandated study of educational growth and opportunity. Department of Education Contract No. LC 89-027-001.

Dauber, S.L., Green, P.J., Hoffer, T.B., Hoogstra, L., Ingels, S.J., Moore, K.A., and Morrison, D. (1996). An Outline of Contextual Measures for the Early Childhood Longitudinal Study. NORC Report to the National Center for Education Statistics under Contract RN94094001.

Dickstein, E. (1972). *The development of self-esteem: Theory and measurement*. Unpublished doctoral dissertation, Johns Hopkins University, Baltimore.

Dreeben, R. & Barr, R. (1988). Classroom composition and the design of instruction. *Sociology of Education 61*, 129-142

Entwisle, D.R., & Alexander, K.L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review, 57*, 72-84.

Entwisle, D.R. & Hayduk, L.A. (1982). *Early schooling*. Baltimore: Johns Hopkins University Press.

Entwisle, D.R., Alexander, K.L., Cadigan, D., and Pallas, A.M. (1987). Kindergarten experience: Cognitive effects or socialization? *American Educational Research Journal, 24* (3), pp. 337-364.

Frazier, J.A. & Morrison, F.J. (1994). Extended-year schooling and growth of language, academic, and general knowledge skills. Manuscript submitted for publication.

Gamoran, A. (1986). Instructional and institutional effects of ability grouping. *Sociology of Education, 59*, 185-198.

Harter, S. & Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development, 55*, 1969-1982.

Henderson, A., Abbott, C., & Strang, W. (1993). *Summary of the bilingual education state educational agency program survey of states' limited English proficient persons and available educational services 1991-92*. Report to the Office of Bilingual Education and Minority Languages Affairs, U.S. Department of Education, submitted by Development Associates, Inc. (Westat, Inc., subcontractor).

Ingels, S.J., Scott, L.A., Rock, D.A., Pollack, J.M. & Rasinski, K.A. (1994). *NELS:88 base year to first follow-up final technical report*. Washington, D.C.: National Center for Education Statistics.

Marcon, R.A. (1994). *Early learning and early identification follow-up study: Transition from the early to the later childhood grades 1990-1993*. Washington, D.C.: District of Columbia Public Schools.

Meisels, S.J., Atkins-Burnett, S., & Nicholson, J. (1995). Assessment of social competence, adaptive behaviors, and approaches to learning. Background paper prepared for the Early Childhood Longitudinal Study, under subcontract to National Opinion Research Center. Revised version published under the same title in the NCES Working Paper Series, Working Paper No. 96-18, August 1996.

Morrison, F.J. (1994). Sources of early literacy: Child, family, and school. Grant application submitted to the Department of Health and Human Services.

Morrison, F.J., Griffith E.M., & Williamson, G.L. (1993). Two strikes from the start: Individual differences in early literacy. Paper presented at the biennial meeting of the Society for Research in Child Development, New Orleans, March.

Mullis, I.V.S., Dossey, J.A., Owen, E.H., & Phillips, G.W. (1993). *NAEP 1992 mathematics report card for the nation and the states*. Washington, D.C.: National Center for Education Statistics.

Pallas, A.M., Entwisle, D.R., Alexander, K.L., & Weinstein, P. (1990). Social structure and the development of self-esteem in young children. *Social Psychology Quarterly, 53*, 302-315.

Puma, M.J., Jones, C.C., Rock, D.A., & Fernandez, R. (1993). *Prospects: The Congressionally mandated study of educational growth and opportunity. The interim report.* Washington, D.C.: U.S. Department of Education.

Rock, D.A. (1994). Results from Prospects. Paper present at the annual meeting of the American Educational Research Association, New Orleans.

Thurlow, M.L., Ysseldyke, J.E., & Silverstein, B. (1993). *Testing accommodations for students with disabilities: A review of the literature.* Minneapolis, MN: National Center on Educational Outcomes.

Ysseldyke, J.E. & Thurlow, M., eds. (1993). *Views on inclusion and testing accommodations for students with disabilities.* Minneapolis, MN: National Center on Educational Outcomes.

2.0 Study Summaries

2.1 Beginning School Study

Purpose of the Study

The Beginning School Study (BSS) was designed to follow a cohort of children as they entered first grade in Baltimore in 1982, and progressed through elementary school. Conceived as a prospective longitudinal study, BSS was designed to examine the social-structural and social and emotional factors that influence achievement in elementary school and beyond.

Four categories of variables are used to predict school achievement and other academic outcomes:

- social-structural (such as race, gender, and socioeconomic status);
- personal (such as cognitive abilities, personal maturity, and special problems);
- social and emotional (such as parents' and children's performance expectations, peer popularity, and self-concept); and
- experiential (such as school marks, attendance, and retention in grade).

These variables are organized as part of a school process model that emphasizes the impact of each set of variables on the following set. Thus, social-structural variables affect personal variables, personal variables may affect social and emotional variables, and all of these may affect achievement. Experiential variables are included to represent the child's experience of his or her environment, and to try to capture the impact of the school's feedback on the student.

Sample Design

A two-stage stratified sampling design was used; schools were selected at the first stage and students were selected at the second. Schools were stratified using information about the racial and social class composition of the student bodies. In all, 20 schools in Baltimore participated in the study. Schools were contacted in 1981-1982 to obtain rosters of kindergarten students so that parental consent could be obtained before the start of school. Approximately 75 percent of the sample was selected from these rosters. The other quarter was selected from rosters of new registrants and first-grade repeaters that were obtained in the Fall of 1982. In all, 825 Baltimore first-grade students were selected to participate in the study. Because over half of the students in Baltimore City Schools were black, white students were intentionally oversampled so that comparisons between the two groups could be supported.

Data were collected from students, teachers, parents, and schools. During the initial two years of the study, when respondents were in first and second grade, two parent interviews, five teacher interviews, and four pupil interviews were conducted. School records, including results of achievement tests, school grades, and absences, were also coded.

Assessment Instruments and Procedures

Cognitive assessments. Baltimore City Schools administered the California Achievement Tests at the beginning and end of each school year and scores from these tests were coded from school records along with information on student attendance and grades. The California Achievement Tests are designed to provide curriculum-sensitive assessments of student achievement. The test was developed in cooperation with teachers and provides scores that are curriculum-referenced as well as norm-referenced. According to the test's developers, 99 percent of all students complete each subtest in the time allotted.

The BSS used Level 11, Form C for first-grade children and Level 12, Form C for second-grade children (Alexander and Entwisle, 1988:32). The first-grade CAT verbal score is a composite of four subtests: phonic analysis, vocabulary, comprehension, and language. The second-grade score includes data from a fifth subtest: structural analysis. The quantitative score combines the results of two subtests: computation and concepts. Correlations between the fall and spring administrations in first grade are .59 verbal and .65 quantitative. In the second grade, the correlations between the fall and spring administrations were .78 for both math and verbal scores.

Currently, several versions of the test are available (CTB/McGraw-Hill, 1994):

- The Basic Skills Battery is comprised of tests in reading, spelling, language, mathematics, and study skills. Both norm-referenced and curriculum-referenced scores are provided. Note that the test designed for kindergarten entry is a readiness test rather than an achievement test and is not scaled to other levels of the CAT.

- The Complete Battery covers all areas from the Basic Skills Battery, plus Science and Social Studies, beginning with first grade. The Complete Battery takes 1.5 hours in kindergarten, 3.5 - 4.5 hours in first grade, and 5+ hours in grades 2 and higher.

- A Survey Form covers the same areas as the Complete Battery, and provides norm-referenced scores in all areas in 2-3 hours. Survey Form tests are not available for kindergarten.

Self-concept. In the spring of first, second, and fourth grades, interviewers administered a brief questionnaire individually to each of the children participating in the study. Included was a 23-item self-concept inventory designed to assess children's self-concept in five domains: character, responsibility, academic, athletic, and appearance. Children responded on a five point scale, ranging from "I am very bad at . . ." to "I am very good at . . .", to the following items:

- Being polite
- Obeying rules
- Being kind
- Being honest
- Being cooperative
- Being helpful
- Being able to look after others
- Being able to take care of yourself
- Learning new things quickly
- Being a good student
- Doing arithmetic
- Reading
- Writing
- Being good at sports
- Playing ball
- Running
- Being strong
- Gymnastics
- Being good looking
- Being just the right weight
- Being just the right height

In an extensive analysis of these data, researchers found that these items did cluster into the five factors listed above. However, among first-grade students the self-esteem factors were highly correlated, with correlations among factors ranging from .59 (character and athletic) to .89 (appearance and responsibility) (Pallas, Entwisle, Alexander and Weinstein, 1990: 306-307). The magnitude of these correlations was smaller in second grade, and smaller still in fourth grade. Pallas and colleagues (1990: 307) interpret this as support for the hypothesis that "facets of the self become more distinct as children mature."

Some group differences in self-esteem were observed. Sizeable gender differences in average levels of esteem were apparent, even in first grade. Girls' estimates of their competence in athletics were much lower than boys. At the same time, girls' ratings in first grade of their academic competence and character were higher than boys' ratings of those same traits. Few black-white differences in self-concept were apparent until the fourth grade. In that grade, black children's estimates of their character, appearance, athletic prowess, and responsibility were higher than white children's estimates of those same traits. Social class differences were not

evident until fourth grade, when advantaged children had higher estimates of their academic competence and character than did their less advantaged peers.

Personal maturity. An inventory of 14 items was adapted from the National Survey of Children (NSC) (Alexander and Entwisle, 1988: 29-30), many of which are also part of the "Behavior Problems Index" used in the National Longitudinal Survey of Youth (NLSY79). In the first wave of the NSC these items were asked of teachers. In this study, homeroom teachers were asked to answer the questions. Teachers were asked whether each item was "exactly like," "very much like," "pretty much like," "somewhat like," "a little like," or "not at all like" each student. Students were rated on the following dimensions:

- very enthusiastic, interested in a lot of things
- fights too much; teases, picks on, or bullies other children
- cannot concentrate, pay attention for long
- usually in a happy mood; very cheerful
- rather high strung, tense, nervous
- not liked by other children
- cheats; tells lies; is deceitful
- shows creativity or originality in schoolwork
- acts too young for age; cries a lot, throws tantrums
- has a very strong temper, loses it easily
- is awfully restless, fidgets all the time, cannot sit still
- keeps to him/herself; tends to withdraw
- very timid, afraid of new things or new situations
- polite, helpful, considerate

Alexander and Entwisle (1988: 29) report alpha reliabilities of .87 and .86 for the first- and second-grade scales, respectively. The correlation between ratings given to children by their first-grade teachers and their second-grade teachers is .48. Subsequent work indicates that these items are represented by three factors: interest-participation, attention span-restlessness, cooperation-compliance (Alexander, Entwisle & Dauber, 1993).

Expectations. Both parents and children were asked what kind of grade they expected the children to receive in school. Parents were asked for their "best guess" for their children's first mark in reading and in mathematics. Responses were coded on a four point scale, ranging from 1 (for unsatisfactory) to 4 (for excellent). These data were collected during the first interview with parents.

Personal interviews were conducted with children during the first quarter of the school year. Interviewers asked children to guess what marks they would receive in reading and in mathematics on their first report card. In the first grade, children were expected to be unfamiliar with grades and the grading process. Consequently, interviewers were trained to explain the report card and grades to the child using a large facsimile of a report card and cardboard cutouts of grades. According to the principal investigators, great care was taken to ensure that children

understood the meaning of grades and report cards. Once interviewers were certain that the child understood the task, the student was asked to guess the grades he or she would get in reading and mathematics by placing the cardboard cut-out grade on the report card.

School record data. Referrals for special services were coded from school records. Referrals to the following ten committees or resource personnel were coded:

- school promotions committee
- school screening committee
- committee on adjustment
- psychological services or individual testing
- speech therapist
- social worker for in-home visitation
- reading resource specialist
- vision or hearing specialist
- physical therapy
- attendance worker

In addition, data concerning attendance, grades, promotion or retention, and achievement test scores were also coded from school records.

Publications and Uses of Data

The BSS has provided data for much important research on the early years of school. Researchers have explored the school achievement model (Alexander and Entwisle, 1988), self-concept (Pallas, Entwisle, Alexander, and Weinstein, 1990), the effects of ability grouping (Pallas, Entwisle, Alexander, and Stluka, 1994), home and school standards for behavior (Alexander, Entwisle, Cadigan, and Pallas, 1987), and many other topics. The major limitation of research based on this study is the study's relatively small sample size. Rather than catalogue the findings from these articles, attention will be focused on the areas in which BSS experience can directly inform the design of ECLS.

Implications for ECLS

Measuring and modeling performance expectations. The BSS is unique among the studies reviewed in that it includes measures of both children's and parents' performance expectations. A strong conceptual framework formed a basis for the design of BSS. In this model, parental expectations concerning their children's performance are thought to affect children's own expectations. In turn, children's expectations are associated with children's actual performance.

Alexander and Entwisle (1988) also point out that expectations are affected by performance. They hypothesize that the grades children receive, an indication of the feedback

given to them by their teacher, temper future behavior and expectations, because children moderate their effort based on their understanding of how that effort is evaluated and rewarded.

Measuring performance expectations among children who have just entered school poses a difficult methodological task. In BSS, interviewers worked closely with first-grade students, explaining the grading process to them. Because the study was conducted within a single school system, the grading scheme was constant for all students, and interviewers could easily answer children's questions. In a national study, this process will be more difficult. Furthermore, while it is typical for first-grade students to receive marks in mathematics and reading, it is less typical for kindergartens to grade students in these subjects. It is unclear what types of performance expectations are appropriate to measure among kindergarten students, or indeed, if the notion of "performance expectation" is a meaningful concept among this age group.

Finally, modeling performance expectations is difficult for two reasons. First, in some sense, parents' and teachers' expectations for grades are proxy measures for the actual grades. Parent and teacher expectations of children's performance have been shown to predict tested student achievement, even when past performance on achievement tests is included as an explanatory variable in the model (Entwisle & Alexander, 1988). This is taken as evidence of the strong impact of expectations on performance. However, the strength of the relationship between teachers' expectations and student performance may result from the fact that teachers' ratings of children's performance is extremely informed, and may in fact be a more reliable measure of student achievement/ability than standardized tests of achievement.

Second, as stated earlier, the relationship between expectations and performance is complex; students' expectations change in response to the information they receive about their performance. Thus, a simple causal model is insufficient. Analyses that incorporate expectations must use sophisticated statistical techniques to estimate the parameters of the model accurately.

School readiness. Alexander and Entwisle (1988) convincingly demonstrate that children's expectations and behavior have a direct impact on achievement. The importance of "personal maturity" in predicting achievement is clear, but the cause underlying this relationship is less clear. It is possible that the behavior problems tapped in this inventory directly interfere with learning; however, this relationship could also be mediated by other factors, one of which may be the "person-environment" fit between the child and classroom. It is possible that students whose behavior does not meet with teachers' standards are less likely to receive direct instruction from the teacher and, consequently, perform at a lower level on achievement tests.

ECLS will explore children's readiness for school and schools' readiness for a diverse group of entering children. The strong link between behavior and performance found in the BSS data calls for further exploration. Kindergarten teachers' standards and expectations for student behavior in the classroom should be measured.

Separating home and school effects. Analysis of data from BSS indicate that the differential cognitive gains observed among advantaged and less advantaged children are a function

of summer experiences, rather than in-school academic experience. In particular, BSS found strong family effects on mathematics growth during the summer months. Students' who were living in families with their fathers gained more during the summer months than those who were living in father-absent families.

A careful disentangling of home and school effects in ECLS will require a design that includes the collection of assessment data at the beginning and end of every school year. Without these two time points, it will not be possible to estimate the effect of schools on students correctly. For example, work conducted by several researchers suggests that the differential achievement gains observed among students from high- and low-socioeconomic status (SES) families results from the fact that high-SES children continue to learn during the summer months, while children from low-SES families do not (Heyns, 1978; Huttenlocher, Levine, and Vevea, under review; Rock, 1994). Further pilot analysis is warranted. The effectiveness of schools for less-advantaged students is an important topic of debate. If ECLS is to contribute to this debate, school effects must be estimated using the appropriate model.

References

Alexander, K.L. & Entwisle, D.R. (1988). Achievement in the first two years of school: Patterns and processes. *Monographs of the Society for Research in Child Development, 53*, Serial No. 218.

Alexander, K.L., Entwisle, D.R. & Dauber, S.L. (1993). 1st grade classroom behavior: its short- and long-term consequences for school performance. *Child Development*, 64, 801-814.

CTB/McGraw-Hill. (1994). *The California Achievement Tests* (CAT). Monterey, CA: CTB-McGraw-Hill.

Dickstein, E. (1972). *The development of self-esteem: Theory and measurement*. Unpublished doctoral dissertation, Johns Hopkins University, Baltimore.

Dreeben, R. & Barr, R. (1988). Classroom composition and the design of instruction. *Sociology of Education 61*, 129-142.

Entwisle, D.R. (1990). Recommendations for the NCES longitudinal study of early childhood/ early education. Paper commissioned by the National Center for Education Statistics.

Entwisle, D.R., & Alexander, K.L. (1988). Factors affecting achievement test scores and marks of black and white first graders. *Elementary School Journal, 88*, 449-471.

Entwisle, D.R. & Hayduk, L.A. (1982). *Early schooling*. Baltimore: Johns Hopkins University Press.

Entwisle, D.R. & Hayduk, L.A. (1988). Lasting effects of elementary school. *Sociology of Education, 61*, 147-159. Baltimore: Johns Hopkins University Press.

Entwisle, D.R. & Hayduk, Leslie (1978). *Too Great Expectations: The Academic Outlook of Young Children*. Baltimore: Johns Hopkins University Press.

Gamoran, A. (1986). Instructional and institutional effects of ability grouping. *Sociology of Education, 59*, 185-198.

Heyns, B. (1978). *Summer Learning and the Effects of Schooling*. New York: Academic Press.

Huttenlocher, J., Levine, S.C., & Vevea, J. (under review). Environmental effects on cognitive growth: The input from school.

Pallas, A.M., Entwisle, D.R., Alexander, K.L., & Stluka, M.F. (1994). Ability-group effects: Instructional, social, or institutional? *Sociology of Education, 67*, 27-46.

Pallas, A.M., Entwisle, D.R., Alexander, K.L., & Weinstein, P. (1990). Social structure and the development of self-esteem in young children. *Social Psychology Quarterly*, *53*, 302-315.

Rock, D.A. (1994). Results from Prospects. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April.

BEST COPY AVAILABLE

2.2 Children of the National Longitudinal Survey of Youth

Purpose of the Study

The National Longitudinal Survey of Youth (NLSY79) is an outgrowth of the National Longitudinal Survey of Labor Market Experience (NLS), a project initiated in the mid-1960s by the U.S. Department of Labor to identify sources of variation in labor-market behavior and training among four groups: men 45 to 59 years of age; women 30 to 44 years of age; and young men and women 14 to 24 years of age. The inclusion of a youth cohort made it possible for researchers to investigate the influence of family factors, school performance, and early work experience on the employment trajectories of young adults. When it was initiated in 1979, the NLSY79 had two primary goals: to replicate many of the analyses based on the earlier cohorts and to help to evaluate the expanded employment and training programs that were available to young people in the late 1970s. The NLSY79 cohort consists of a national sample of civilian and military youth (male and female) who were between the ages of 14 and 21 in 1979. Hispanics, blacks, and economically disadvantaged whites were intentionally overrepresented in the sample. When the NLSY79 was initiated, there were fewer Asian Americans residing in the United States than at present. Consequently, Asian Americans and other groups that have become more numerous since 1979 are not among those represented in the supplemental sample.

A new NLS youth cohort, the NLSY97, is being surveyed in 1997, with follow-ups scheduled for the years thereafter. The NLSY97 sample comprises youth between ages 12 and 16 as of December 31, 1996.

The NLSY79 includes a core set of questions on the following topics: marital history, schooling, current labor force status, jobs and employer information, gaps in employment, training, work experience and attitudes, military service, health limitations, fertility, income and assets, household composition, and geographic residence. Information on each of these topics has been collected yearly since 1979. In 1982, the National Institute of Child Health and Human Development (NICHD) provided funding to incorporate a comprehensive set of questions on fertility and childcare into the NLSY79. These questions were included in the yearly administrations of the NLSY79 from 1982 to 1986 and were asked again in 1988, 1990, and 1992. In 1986, with funding from NICHD, the NLSY79 was again expanded to include a battery of cognitive, socioemotional, and physiological assessments of the children of female respondents, as well as an assessment of the quality of the children's home environments. These assessments have been administered biennially since 1986. The inclusion of these assessments allows researchers to examine how a variety of maternal and familial characteristics and behaviors affect children.

Sample Design

The NLSY79 sample is a multistage, stratified random sample. The sample was identified through a random selection of counties, an enumeration of district-block groups, and a subsequent screening of 75,000 households. The original NLSY79 sample comprised the following groups:

(1) a cross-sectional sample of 6,111 civilian youths between the ages of 14 and 21 in 1979; (2) a supplemental sample of 5,295 Hispanic, black, and economically disadvantaged white youths; and (3) a sample of 1,280 youths who were from 17 to 21 years of age in 1979 and who were enlisted in one of the four branches of the military. The military sample continued only for the first five years; all but a subsample of 201 randomly selected youths from the military sample were dropped from the study after 1984. Beginning with the 1991 survey year, economically disadvantaged white respondents from the supplemental sample were also dropped from the study. Response rates for respondents who have remained eligible to be interviewed have remained at or over 90 percent for the first twelve years of the study.

The children's sample of the NLSY79 consists of all children born to female civilians in the NLSY79 cohort who reside with their mothers during a given survey year; children of male respondents are not included in the sample. Of the 5,828 civilian women who were in the original NLSY79 sample, 3,053 were identified as having children by the 1986 survey round.

Approximately 5,000 children (96 percent of those identified) were included in the 1986 Child Survey; children born since 1986 have been included in later survey rounds. The children's sample of the NLSY79 does not constitute a nationally representative sample of children. In 1986, female respondents with children were between the ages 21 and 28; their children ranged in age from less than 1 year to about 15 years. Consequently, the majority of children surveyed were born to teenage or young adult mothers, a group that would in all likelihood be less educated and otherwise of lower average socioeconomic status than a full cross-section of mothers.

Assessment Instruments and Procedures

The NLSY79 Child Survey is a set of assessments that is designed to measure children's socioemotional and physiological well-being as well as their performance on certain verbal, mathematical, and memory tasks. The Child Survey consists of two survey schedules, the Mother Supplement and the Child Supplement.

The Mother Supplement. The Mother Supplement is designed to be filled out by the mother during the interviewer's administration of the Child Supplement. Interviewers are instructed to provide assistance to any respondent who has difficulty completing the Mother Supplement. A Spanish version of the supplement is made available to respondents whose primary language is Spanish. The Mother Supplement contains the following sections:

> **The HOME Short Form:** includes items from the Home Observation for Measurement of the Environment, an inventory developed by Caldwell and Bradley (1984) that contains age-specific versions of a set of scales designed to measure the quality of cognitive stimulation and emotional support provided to children by their families.
>
> **How My Child Usually Acts (temperament):** (for children under 7 years of age) includes items from Rothbart's Infant Behavior

Questionnaire and Campos and Kagan's Compliance Scale that together form a set of maternal report scales measuring temperament or behavioral style over the previous two-week period.

Motor and Social Development: (for children under 4 years of age) includes items drawn by Dr. Gail Poe from Bayley (1969), Gesell (1947), and the Denver Developmental Screening Test that measure various milestones in the areas of motor, cognitive, and social development for young children.

Behavior Problems Index: (for children 4 years or older) includes items from Zill and Peterson's special adaptation for the NLSY Child Study (see Baker, Keck, Mott & Quinlan, 1993, p.105) of the Child Behavior Checklist developed by Achenbach and Edelbrock (1981, 1983), that elicit maternal ratings of various areas of problem behavior exhibited by children (e.g., hyperactivity, anxiety, dependency, depression, aggressiveness).

School and family background: (for children 10 years or older) information on schooling, grade repetition, school behavior and expectations, peer relations, and religious attendance and training.

The Child Supplement. The Child Supplement is used by the interviewer to collect (1) general and health-related information from the mother for each child; (2) responses from the child on nine assessment instruments; (3) interviewer evaluations of the child's attitudes toward testing; and (4) interviewer observations of the quality of the child's home environment. Spanish versions are available for most of the assessments contained in the Child Supplement. The Child Supplement contains the following sections:

Background information: includes identifying information (age, gender, grade in school) obtained from the mother for each child. Information on the child is linked to identifying information for the mother (obtained in the main NLSY79 survey) through the child's ID number. This number is assigned to children during the administration of the main NLSY79 Survey.

Child's health profile: includes information from the mother on the child's health limitations, accidents and injuries, medical treatment in the last twelve months, health insurance coverage, and measures of the child's height and weight at the time of the interview.

Parts of the Body: (for children 1 to 2 years of age) includes ten items, developed by Kagan (see Baker, Keck, Mott & Quinlan, 1993, 113-116), that measure young children's ability to identify various parts

of their body. This assessment was used in 1986 and 1988 but not in subsequent survey years.

Memory for Location: (for children 8 months to 3 years of age) developed by Kagan (Kagan, 1981) to measure young children's ability to remember the location of an object subsequently hidden from view. This assessment was used in 1986 and 1988 but not in subsequent survey years.

McCarthy Verbal Memory Scale: (for children 3 to 6 years of age) a subscale of the McCarthy Scales that assesses children's short-term verbal memory (i.e., their ability to remember words, sentences, or major concepts from a story). The story segment of the assessment was removed from data collection after the 1990 survey.

What I Am Like/SPPC: (for children 8 years or older) two scales from Harter's Self-Perception Profile for Children that measure perceived self-competence in the academic domain and sense of general self-worth.

Memory for Digit Span: (for children 7 years or older) a component of the revised Wechsler Intelligence Scales for Children that assesses the ability of children to remember and repeat numbers sequentially in forward and reverse order.

The Peabody Individual Achievement Test (PIAT) Mathematics subtest: (for children with a PPVT age of 5 years or older) a wide-range measure of achievement in mathematics for children. An adaptation of the administration form in the Child Supplement is accompanied by the standard PIAT materials contained in Volumes I and II of the PIAT Easel Kit.

The PIAT Reading Recognition and Reading Comprehension subtests: (for children with a PPVT age of 5 years or older) assesses the attained reading knowledge of children. The item format in the Child Supplement replaces the standard PIAT record booklet but interviewers use the official item plates and instructions for administration contained in Volumes I and II of the PIAT Easel Kit.

The Peabody Picture and Vocabulary Test-Revised (PPVT-R Form L): (for children with a PPVT age of 3 years or older) a measure of children's receptive vocabulary. As with PIAT, children are shown the official item plates and their responses are recorded in the Child Supplement.

Interviewer evaluation of testing conditions: gauges the attitudes of the child toward testing, the child's general physical condition, and records any events that may have interfered with or caused the premature termination of an assessment.

Interviewer observations of the home environment: a subset of the HOME-Short Form items that indicate the interviewer's perceptions of mother-child interaction and the nature of the physical environment. The remaining items of the HOME-SF are maternal report items and are contained in the Mother Supplement.

The Child Self-Administered Supplement (CSAS): (for children 10 years or older) a self-report booklet filled out by older children that was first developed for the 1988 interviews. The CSAS collects information on a wide range of topics, including parent-child interactions, attitudes toward school, extracurricular activities, peer relationships and dating, and alcohol consumption and drug use. The contents of the booklet have been expanded since 1988 and now include information on any children that are born to NLSY79 children age 13 and older.

Administration and timing of the child assessments. Data collection for the NLSY79 child sample is carried out using personal home interviews and occurs in conjunction with the main NLSY79 interviews (conducted with the child's mother). The main interviews with mothers are conducted yearly; the child surveys are carried out every two years. According to the authors of the NLSY Child Handbook (Baker, Keck, Mott & Quinlan, 1993), the main NLSY79 interview takes about an hour to complete; in their estimation, the Child Survey (both Mother and Child Supplements) adds about 30 minutes to the total survey time. Although interviewers generally try to schedule the Mother and Child Supplements on the same day, the supplements are sometimes completed during separate visits to the home. Not all components of the Child Assessments are administered during every survey year. Some assessments are completed only once by a child--Verbal Memory (age 4 to 6), Digit Span (age 7 and over), and PPVT-R (age 4 and over)--the first time they became age-eligible. With the exception of Parts of the Body and Memory for Location (dropped after 1988) and the story segment of the McCarthy Verbal Memory Scale (dropped after 1990), the other assessments have been completed during each survey year for all age-eligible children. Starting in 1988, children who were 10 to 11 years of age were administered all assessments for which they were age-eligible regardless of which assessments they had previously completed. The data collection for older children was expanded in 1988 to provide researchers with more complete data on a group of children who would ultimately (after several survey rounds) serve as a large, more fully representative sample of early adolescent youth.

Publications and Uses of the Data

The emphasis in the NLSY79 data on labor force behavior, ethnic diversity, extent of impoverishment, and the relative youth of mothers has allowed researchers to investigate the effects on children of specific demographic and social phenomena. Several studies using the NLSY79 child assessment data have examined the effects on children of maternal employment, use of child care, adolescent parenthood, divorce and father absence, multigenerational parenting, and poverty. Chase-Lansdale, Mott, Brooks-Gunn, and Phillips (1991) cite many of the recent studies on these topics and note several important issues that can be addressed by the NLSY79 child data.

Maternal employment. NLSY79 data allow researchers to explore the timing of mother's employment, particularly during or after the child's infancy and to examine the effects of early return to work on children's socioemotional and cognitive development. A number of studies have begun to explore the timing of employment (see Chase-Lansdale et al., 1991, for references). A common finding among these studies is that maternal employment in the first year of life has significant long-term effects on both cognitive and socioemotional development; studies differ on which groups are most affected. Divergent findings on the specific groups affected may help to promote refinement of the original question. Chase-Lansdale et al. note that one question that needs to be pursued is the role of economic context in maternal employment.

Child care. In both 1986 and 1988, but not 1990, child-care questions were asked of both employed and nonemployed mothers regarding use of child care during the prior four weeks. In these same years, mothers were also asked to provide retrospective data on child care use during the first, second, and third years of life. The availability of data on different types of primary- and secondary-care arrangements offer several advantages in studying their effects on children: (1) it enables analysis of child outcomes in naturally-occurring contexts, including both formal and informal arrangements, an option unavailable to studies that have focused on center care; (2) retrospective data on past types of care can be used to examine the differential effects of type of child care, for example, Baydar and Brooks-Gunn (1991) found that grandmother care is beneficial especially to economically disadvantaged white children, as reflected in higher PPVT-R scores and fewer behavioral problems; (3) NLSY79 data facilitate an analysis of child care experience over time within specified patterns of maternal employment; and (4) NLSY79 data allow examination of family and child characteristics as potential moderators of children's experience with child care (e.g., family income, education, stress level, and child's gender and temperament). Limitations of the child care data are also noted: (1) information on current child-care use is restricted to a narrow 4-week period; (2) no measures of the quality of child care are included; (3) respondent burden in a survey of NLSY79's scope necessitates a reduction in detail for some topics (e.g., data on child-care costs reflects costs for all arrangements rather than different arrangements by child); (4) the child care questions were eliminated in the 1990 survey round as a cost saving measure; (5) it is difficult to assess the reliability of retrospective reports on patterns of child care.

Adolescent pregnancy and parenthood. The substantial number of young mothers in the NLSY79 make possible studies of various aspects of adolescent (and usually single) parenthood.

Most studies of effects of teenage parents on childhood have focused on infant outcomes. The NLSY79 dataset is advantageous first because it allows researchers to address the effects of teenage parenting on preschool and school-age children (the effects on infants cannot be studied using NLSY79 because the first child assessments occurred when mothers were 21 to 28 years old). Second, it enables researchers to disentangle the impact of maternal age from other factors associated with adolescent parenthood (e.g., economic disadvantage, low educational attainment, mother-headed families). Because of the large number of economically disadvantaged families represented in the study, it is possible to construct comparison groups of older mothers who have individual and family backgrounds similar to teenage mothers and thereby isolate the effects of maternal age from other factors affecting child development. Extensive data on the school, work, and individual histories (e.g., histories of cigarette, alcohol, and drug use) of NLSY79 mothers may help to identify the factors that lead to early motherhood and in turn have negative consequences for the child. Third, the NLSY79 permits researchers to examine the relationship between the life-course trajectories of mothers and children; changes in the mother's life trajectory (e.g., leaving welfare to enter the workforce) can have significant effects on the trajectories of children. NLSY79 permits researchers to examine the relationship of interlocking trajectories at multiple points in time. Limitations are also noted: (1) researchers are unable to examine the relationship for the period from infancy to early adolescence for the oldest children in the study because the first data collection for children occurred when some were already in their teens; and (2) no information on the youngest mothers is available prior to 1979, when they were 14.

Divorce and nonmarital childrearing. The NLSY79 has a higher incidence of divorce than nationally representative samples because of the overrepresentation of adolescent mothers and economically disadvantaged groups. The advantages offered to investigators of divorce and marital disruption include: an opportunity to study divorce among Hispanics and blacks, a phenomenon that is understudied in these groups; prospective studies of the effects of divorce on children, as well as an opportunity to disentangle the conditions existing before divorce from the sequelae of divorce; (because of the longitudinal design) examination of the mediating effects of child age on the impact of divorce; and investigation of the impact of father visitation and the presence of father substitutes, because a large number of children in the sample have never lived with their fathers.

Children in poverty. Because of the large number of economically disadvantaged families included in the sample, NLSY79 offers a number of advantages in studying the effects of poverty on children: (1) the data set permits researchers to disentangle many of the factors that are associated with poverty (e.g., low maternal education, poor schools, father absence, and minority status) and to examine their separate effects on children; (2) it allows researchers to compare neighborhoods in terms of poverty concentrations and to examine the effects on children of living in neighborhoods thus categorized; (3) it allows researchers to document patterns of persistent poverty because of the availability of extensive income and employment histories for NLSY79 participants; (4) it also allows researchers to investigate the extent to which poverty is mediated by access to various transfer payments and the effects of these programs for children.

Multigenerational parenting. Because the NLSY79 sample includes the large numbers of single mothers and mothers who work and because of the existence of data on household composition and residence history, investigators can document the incidence of multigenerational households and examine the points at which mothers' and children's life cycles and multigenerational living arrangements intersect. The NLSY79 data set may also be of use in controlling for environmental and sociodemographic characteristics of families when examining the effects of social structure on children. The existence of a comprehensive residential history for the first 18 years of mothers' lives can help to clarify the extent to which complex family structures have a cross-generational history, with parental patterns repeating themselves from one generation to the next.

Implications for ECLS

Several of the assessment instruments used for the NLSY79 Child Survey seem plausible candidates for inclusion in ECLS, or to hold lessons for the ECLS assessment methodology. Some of the cognitive assessments may provide source material for the ECLS assessment batteries. The method of cognitive assessment administration in the NLSY79 Child Survey -- one-on-one, with the assistance of assessment easels and a laptop computer -- is the same as that envisioned for kindergarten and first grade in the ECLS. Since no other national large scale assessment utilizes this methodology (though more local studies such as the Greensboro Early Schooling Study conduct one-on-one assessments without the use of computer technology), there are potential lessons for the ECLS in terms of adaptation of computer technology to assessment administration and scoring. There are also potential lessons in terms of administration procedures, assessor training, and assessment evaluation and quality control. Although the NLSY79 Child Supplement is household-based and the ECLS will be school-based, both are similar in that both conduct (or will conduct) assessments in many dozens of sites across the nation and involve hundreds of assessors and the need for a high quality, highly standardized approach. In addition to implications for the cognitive assessments, three of the remaining assessments used with children 5 years or older may hold promise or contain lessons for the ECLS: (1) those measuring the nature and quality of parent-child interactions, (2) children's behavioral problems, and (3) the child's self-concept. Below we discuss the cognitive measures, then the importance, strength and weaknesses of the measures of parent-child interaction, behavioral problems, and child self-concept.

Cognitive Assessments: vocabulary, reading, mathematics. The NLSY79 Child Survey included such cognitive assessments as a vocabulary test (the PPVT-R), a reading test (the PIAT), and a math test (the PIAT). (The Greensboro Early Schooling Study employs these same instruments, although it utilizes a more recent version of the PIAT [PIAT-R].) The three tests overlap with two of the three areas provisionally identified for study in the ECLS, language and mathematics. Moreover, portions of the PIAT not used by NLSY79 (but employed in the Greensboro study), specifically, the General Information subtest, extend its range to the third area of interest for the ECLS, general knowledge, comprising understanding of the physical, biological and social world (or, roughly, science and social studies). There are, moreover, a number of other off-the-shelf tests that in format and function resemble the measures used on NLSY79 and

that can be used to measure the cognitive status and growth of young children in the areas of mathematics, language, and general knowledge. (These other measures include: the Woodcock Johnson Psychoeducational Battery-Revised Edition, Primary Test of Cognitive Skills [PTCS], Children's Cognitive Battery [CCB], Test of Early Mathematical Abilities - 2nd Edition [TEMA-2], and the Test of Early Reading Abilities - 2nd Edition [TERA-2]) should To what degree could such measures be adapted to the use of the ECLS?

In order to make judgments about the utility of such measures for ECLS, it will be useful to summarize some of the key objectives and criteria guiding assessment development for the study.

First, the ECLS assessment batteries must be sensitive to the curriculum, that is, to what is being learned in kindergarten at a particular point in time (specifically, 1998-99), and to later grades thereafter. To develop a battery that reflects curricular content and objectives in the areas of reading, mathematics and general knowledge, reviews will be conducted of various standards documents, textbooks, work books and curriculum guides. Because general cognitive assessments such as the PIAT were not designed specifically to measure school learning over a six year span nor to reflect curricular trends that will be dominant at the end of the century, it is doubtful that any existing assessment could, by itself, fully meet the purposes of the ECLS. Nonetheless, the tests used in the NLSY79 -- the PIAT and the PPVT-R -- and a number of other off-the-shelf tests -- supply a store of powerful items with well documented psychometric characteristics that could contribute substantially to an overall item pool for the ECLS.

To be included in the ECLS item pool, any item, from whatever source, must be related to the domain content and skills outlined in the ECLS assessment framework, which has been based on a review of the school curriculum; must provide reliable, valid, unbiased measurement for subpopulations as well as the population at large; and must provide a basis for measuring cognitive growth over time. In addition, at least a substantial subset of items in the ECLS pool must be suitable for use in group administration, since one must link between the early rounds of individual administration (kindergarten and first grade) and later rounds that may be subject to group administration for most students (grades two through five). Many items from existing measures will meet these several criteria. In short, while off-the-shelf tests will not meet the full purposes of the ECLS, many of the items within existing batteries are likely to be strong candidates for inclusion within the battery. The final ECLS battery is likely to be a combination of proven, existing items, and new items, specially written for the ECLS.

As indicated above, other promising NLSY79 measures for children 5 years or above include: measures of the nature and quality of parent-child interactions, of children's behavioral problems, and of the child's self-concept. The importance of these measures and some of their strengths and weaknesses are addressed below.

Parent-child interaction. Since parents' daily interactions with children undoubtedly influence children's school achievement and shape children's self-concepts as academic performers, some measure of the nature and quality of parent-child interactions seems warranted.

Parents' beliefs about their children's abilities and their expectations regarding their children's academic performance have been shown to be a factor in children's school achievement and may be particularly important in the early years of school (Alexander and Entwisle, 1988; Entwisle and Hayduk, 1982; Hess, Holloway, Dickson, and Price, 1984; Stevenson and Newman, 1986).

The NLSY79 child survey primarily uses the Home Observation for Measurement of the Environment-Short Form (HOME-SF) to measure the nature and quality of parent-child interactions. The HOME-SF includes items from Caldwell and Bradley's HOME Inventory (1984), a set of observational measures of the quality of cognitive stimulation and emotional support provided to children by their families. The original inventory was designed to be filled out by an outside observer. As its name suggests, the Short Form is an abbreviated version of the original inventory and has been modified for survey research. The majority of items consist of multiple-response maternal reports; the remaining items consist of interviewer perceptions of the physical environment and the quality of mother-child interactions. The HOME-SF is divided into four parts, each tailored to a particular age group: 1) children under 3; 2) children between the ages of 3 and 5; 3) children between the ages of 6 and 10; and 4) children 10 years and older.

Several maternal-report items are designed to measure the cognitive stimulation provided to children by their families. For 3- to 5-year-olds these items include the frequency with which stories are read to the child; the number of books in the child's possession; the child's (individual or shared) possession of least five children's tapes or records; help provided to the child at home in learning the alphabet, numbers, shapes, and colors; the frequency with which the child is taken on family outings (e.g., shopping, to the park, on a picnic); and the frequency with which a family member has taken or arranged for the child to be taken on trips to a museum (e.g., children's, art, historical, or scientific museum). Five interviewer observations are also included as measures of cognitive stimulation for 3- to 5-year-olds, which focus on aspects of the child's physical environment (safety of play environment and presence or absence of structural or health hazards in the home), quality of the perceptual environment (e.g., darkness or perceptual monotony of interior rooms), and the relative cleanliness and clutter of visible areas.

A smaller number of maternal-report items are included as measures of emotional support provided to children. For 3- to 5-year-olds these include the extent to which the child is allowed to choose what foods he/she eats for breakfast and lunch; the number of hours the television is on in the home each day; the mildness or severity of the caregiver's response to the child's expressions of anger (e.g., hitting the mother or guardian); the frequency with which the child eats a meal with both the mother and father (or stepfather/father figure); and the number of times the child has been spanked in the previous week. Several interviewer observations of mother-child interactions are also used to assess the quality of the emotional relationship between caregiver and child, including observations regarding the extent to which the caregiver conversed with the child and verbally answered the child's questions or requests; affectionate displays by the parent toward the child (whether through physical displays such as hugging or through vocal displays of praise or positive feeling); and displays of anger or displeasure such as shaking, grabbing, slapping, or spanking the child.

The items used for assessing the home environments of older children (6 to 10 years or 10 years and older) include many of these same items with age-appropriate modifications. Age-appropriate items are also added for older children, and include maternal reports on the child's responsibilities around the house (e.g., making the bed, cleaning one's room, doing routine chores, getting up on time, getting ready for school); parental encouragement of the child's involvement in hobbies or activities (e.g., music, dance, sports, including opportunities to take special lessons); and the parent's likely response to low grades on a report card (e.g., contacting the teacher or principal; punishing the child; helping the child with schoolwork).

The items measuring both cognitive stimulation and emotional support cover a wide range of age-appropriate activities and emotional displays and appear to be valid measures of the quality of parent-child interactions. Reports on the predictive validity of the HOME and HOME-SF confirm its usefulness as a predictive measure. Several studies have shown that both the original inventory and the HOME-SF predict later cognitive and social development (see Baker et al., 1993 for references). The construct validity of the HOME-SF has also been confirmed by a number of studies. Using exploratory and confirmatory factor analysis, Parcel and Menaghan (1989) demonstrated that the 1986 HOME-SF data generated conceptually similar scales to those developed by Bradley and Caldwell for the original HOME inventory. Menaghan and Parcel (1992) found that three scales—cognitive stimulation, maternal responsiveness, and good physical environment—were reliable and stable across time for preschool and school-age children. These scales were also found to correlate with expected social dimensions such as parental education, SES, race, and marital status.

Although the HOME-SF includes both maternal reports and interviewer observations, it does not consistently include measures from both observers on all aspects of the home environment. For example, interviewers do not provide estimates of the number of books, magazines, or other material resources available for use by children. In addition, no interobserver reliabilities have been calculated for the HOME-SF. Since parents might be inclined to give answers that they believed were socially acceptable (e.g., over-reporting time spent with children in certain activities), the absence of interobserver reliabilities leaves the reliability of the HOME-SF measures open to question.

A second problem with the HOME-SF scale has to do with its focus on the material resources available to children as measures of cognitive stimulation. The scale provides no measures of the parents' attitudes toward learning or their specific expectations regarding their children's school performance. Such attitudes and expectations are conceivably as important to children's school achievement as the material resources needed to engage in academic pursuits. The inclusion of items regarding parents' beliefs, attitudes, and expectations regarding learning and school performance could also help to resolve some the problems posed by the absence of interobserver reliabilities (e.g., if items from the HOME-SF were administered via telephone interviews with parents). If parents' responses were presented within the context of their own beliefs about the importance of particular activities in preparing children for school, they might be less inclined to provide answers thought to be acceptable to the interviewer. The consistency of parents' responses to questions regarding the importance of certain activities and the frequency

with which they engage in those activities with their children could also be evaluated. Although a few items are included for children 6 years or older regarding parents' involvement in children's schooling, these items are absent in the questionnaires for younger children. Since many children between the ages of 3 and 6 are in preschool or kindergarten these questions should also be asked of parents with younger children. Ideally, an independent measure of parents' involvement with schooling (e.g., attendance at parent-teacher conferences, interest expressed in their child's academic progress or difficulties) should be obtained from children's teachers.

The emotional support items in the HOME-SF rely heavily on interviewer observations (7 of 12 items for 3- to 5-year-olds). If ECLS interviews parents by phone, these measures could not be used. However, other questions might be included that address parents' beliefs about disciplining children and that identify behaviors that typically result in praise or disciplinary action. Questions regarding types and frequencies of particular rewards and punishments could be formulated within this context. Questions addressing children's perceptions of their parents' acceptance of particular attitudes and behaviors could be included to provide a more balanced design. Harter (1985; Harter and Pike, 1984) has developed a set of social acceptance scales for use with young children that might be used for ECLS (see below).

Behavioral problems. Alexander and Entwisle observe that "the linkage between socioemotional status and school performance may be the route by which young children most affect the course of their own development" (1988: 104). For children who are just beginning school, behavioral dispositions may be particularly important because they can shape teacher and peer evaluations in ways that affect academic progress. Teachers may give more positive evaluations to children who work well with other students and who appear to be happy and well-adjusted; peers may be more willing to include such children in group activities. Children who are easily distracted or who disrupt classroom activities may in turn be less likely to benefit from classroom instruction.

The Behavioral Problems Index (BPI) is the primary measure of children's behavioral dispositions used in the NLSY79 child survey. The Behavioral Problem Index includes 28 maternal-report items that assess problem behaviors along six dimensions: antisocial, anxious-depressed, headstrong, hyperactive, immature dependency, and peer conflict-social withdrawal. Mothers are asked to report on specific behaviors exhibited by the child in the previous three months. For each of the items included, respondents are instructed to indicate whether a statement is often true, sometimes true, or never true for the child (e.g., "He/she is restless or overly active, can't sit still").

The BPI was developed for children ages 4 to 17 by Peterson and Zill (cf. Peterson and Zill, 1986) and includes items from Achenbach and Edelbrock's (1981) Child Behavior Checklist as well as other child behavior checklists (Graham and Rutter, 1968; Kellam, Branch, Agrawal, and Ensminger, 1975; Rutter, Tizard, and Whitmore, 1970). The items used in the NLSY79 were developed from a larger set of items originally used in the 1981 Child Health Supplement of the National Health Interview Survey. The BPI has also been used in the National Survey of Children

a similar set of items was also included in the Baltimore Beginning School Study (cf. Alexander and Entwisle, 1988).

The BPI has successfully discriminated between clinic and nonreferred children in the National Child Health Supplement (Zill and Snyder, 1981), between children from high-conflict marriages and those from low-conflict marriages (Peterson and Zill, 1986), and between children of divorced and remarried parents and those from nondivorced families (Zill, 1988). In studies based on the NLSY79 child data, higher scores on the BPI (indicating higher levels of problem behaviors) have been linked with family poverty, divorce, and father absence. Mothers who are younger and less educated also report a higher incidence of problem behaviors for their children. Child characteristics have also been linked with BPI scores. Older children tend to score higher on all areas of the BPI; children who have been referred for psychological help also have higher scores. The Behavioral Problems Index thus appears to be both a valid and a useful measure of children's problem behaviors. The modest internal reliabilities reported for subscales (r's range from .54 to .71) argue against their isolated use. The overall scale, however, has proven to be highly reliable (reported r's range from .86 and .92) (Baker et al., 1993; Chase-Lansdale et al., 1991; see Baker et al., 1993 for procedures used to define the subscales and to verify their internal consistency). For NLSY79, only mothers were asked to assess children's problem behaviors, but ideally, assessments of children's behavior should be obtained from both parents and teachers.

Evaluations from teachers as well as parents may reveal differences in the contexts (home vs. school) in which problem behaviors are exhibited. Divergent standards and criteria may also be used by parents and teachers in assessing children's behavior. Consequently, questions regarding respondents' standards for appropriate behavior and their typical responses to problem behaviors should be included in parent and teacher interviews. Evaluations of children's responses to feedback or criticism might also be included to more fully assess children's socioemotional development; similar questions could also be asked of children (see below, regarding Harter's behavioral competence scale for children). Questions regarding specific events (e.g., death of a family member, a move to a new school or residence, a recent illness) that might have triggered particular behaviors would also be useful to evaluate such behaviors. The inclusion of items that assess more positive aspects of child behavior might also be considered for ECLS. A few items assessing positive behaviors were included for balance in the scales used for the BSS (cf. Alexander and Entwisle, 1988). Since some parents and teachers might be reluctant to focus only on a child's problem behaviors, a more balanced mix of positive and negative items might be desirable.

The child's self-concept. Children's perceptions of their own competence can play an important role in their success in school. While parents, teachers, and peers undoubtedly influence children's perceptions of their own competence, the ways in which children respond to the evaluations of others, and the extent to which they value or discount their judgments, can serve to moderate the effects of such influences. Children's sense of acceptance by parents, teachers, and classmates can also have important consequences for academic achievement. Children who feel excluded from group activities or who feel disliked by a teacher may be less likely to participate in activities that offer opportunities to develop new skills.

The NLSY79 child survey uses two subscales of Harter's Self-Perception Profile for Children (SPPC) to measure children's perceived competence in the academic skills domain and their sense of global self-worth. The scales are used only with children who are 8 years or older; age-appropriate versions of the academic competence scale are available for younger children. As Harter (1985) notes, the global self-worth scale should not be used with young children because children do not develop a consolidated concept of global self-worth until middle childhood.

The Pictorial Scale of Perceived Competence and Social Acceptance for Young Children (Harter and Pike, 1984) was developed for use with children whose reading skills and understanding of trait labels (e.g., smart, popular, good-looking) were insufficient to complete the Self-Perception Profile for Children. Two versions of the Pictorial Scale are available, one for preschool and kindergarten children, and one for first and second graders. Subscales include Cognitive Competence, Physical Competence, Behavioral Competence, and Peer, Maternal, Father, and Teacher Acceptance. Each subscale includes six items. The Behavioral Competence subscale and the Father and Teacher Acceptance subscales were developed for a revised version of the Pictorial Scale and have not been widely used. The Self-Perception Profile, used with older children, comprises a similar set of scales but also includes the global self-worth scale and a physical appearance scale. Teacher rating scales have been developed that parallel the items administered to children and are used to elicit teachers' ratings of a child's cognitive competence, physical competence, and peer acceptance.

The pictorial format used with younger children offers several advantages, because it is more likely to engage the child's interest than a verbal interview or questionnaire and thus to sustain the child's attention to the task. The pictorial format also permits the concrete depiction of specific skills and activities. The graphic presentation of activities such as puzzle-solving, playing with friends, or riding a bike avoids the problem of verbal descriptions of skills and traits (e.g., smart, popular, athletic) that may be beyond the comprehension of young children. The different versions of the Pictorial Scale that are available for preschool and kindergarten children and for first and second graders reflect the different criteria for academic and social competence for these age groups. Harter and Pike (1984) note that "puzzles may be indicative of cognitive competence during the preschool and kindergarten years, but more scholastically oriented skills such as being able to spell, read, or add are better measures of cognitive competence in the first and second grades" (p. 1970).

Factor analysis yielded moderate to high loadings on designated factors (competence and acceptance) for preschool-kindergarten and first-second grade samples for four of the subscales (cognitive competence, physical competence, peer acceptance and maternal acceptance); loadings were somewhat higher for the first and second grade samples (Harter and Pike, 1984). Subscale reliabilities for both age groups ranged from .52 to .85; the acceptance subscales were found to be somewhat more reliable than the competence subscales. Harter and Pike (1984) note that the lower reliability of the competence scales is attributable to high item means for these scales; i.e., most children perceived themselves to be competent at the tasks that were listed (e.g., good at counting, good at the alphabet). One implication of these findings is that the scales may not be adequately calibrated to the age-groups being tested. For example, reliabilities on the competence

BEST COPY AVAILABLE

subscale are .71 for preschoolers but only .52 for kindergartners, suggesting that different items may need to be used for these two age groups or that additional items may be needed to tap more specific skills within this domain. Reliabilities for the overall scale (comprising in this instance the cognitive and physical competence scales, and the peer and maternal acceptance scales) are reasonably high (r's range from the mid to high .80s).

When asked to give reasons for their perceived competence on a specific task (e.g., "How do you know you are good at counting?"), first and second-graders were able to provide reasons that were consistent with their assessments, suggesting that the ratings are valid in the sense that children's perceptions of competence are based on specific behavioral referents. No systematic data are available for younger children, but available evidence suggests that younger children are also able to justify their responses.

The subscales used with first and second graders successfully discriminated between groups of children predicted to differ in each domain. The cognitive subscale discriminated between first graders who were promoted and those were held back; the peer acceptance scale discriminated between children who were new to a particular school from those who had attended the school for a year or longer; and the physical subscale discriminated between children who had been preterm infants and those who had been fullterm infants.

The predictive validity of the scales is not well established. However, Bierer (1981) found that, for children who overrated their competence relative to their teachers, perceived competence was not predictive of the children's behavior. Such children chose tasks that were congruent with their actual rather than their perceived competence (e.g., they chose puzzles that were relatively easy to solve).

The findings reported by Harter and Pike (1984) are based on small samples of children between the ages of 4 and 7 from middle-class schools. No large-scale testing has been done with children of this age. As noted, two subscales of the Self-Perception Profile for Children (SPPC) have been used in NLSY79 with older children. Internal consistency reliabilities for the 1990 administration of the NLSY79 were .69 for the scholastic competence subscale and .67 for the global-worth subscale. Fairly strong associations were found between perceived scholastic competence and performance on the PIAT assessments; associations between global self-worth and PIAT scores are much weaker. As might be expected, both within and cross-year correlations between perceived scholastic competence and performance on various assessments is more pronounced for older children.

Harter (1985) notes that the SPPC is not appropriate for use with special populations such as mentally retarded or learning disabled children. Special versions of the scale have been constructed for these groups. The scale may also need to be modified for other groups. She also observes that children's scores on the SPPC are affected by the particular reference groups they employ. Individual interviews with mainstreamed retarded children revealed that the children compared their performance to that of other mentally retarded children and consequently did not regard their performance as deficient. In contrast, mainstreamed learning disabled children

compared themselves with regular classroom children and considered themselves to be less scholastically competent than their classmates. Harter urges that information be obtained on the social comparison groups employed, especially in dealing with special populations. Questions regarding the basis for children's self-judgments can be helpful in evaluating their perceptions of competence in particular domains. Some children may base their judgments on comparisons to particular reference groups, others may rely on feedback from parents and teachers, still others may use performance or behavioral criteria (e.g., I'm smart because I get my homework done in class). Children may also have different perceptions about how they developed particular skills or how they might improve in areas where their skills are weak. Questions addressing these issues may be particularly important in designing interventions to help children improve particular skills. With regard to children's global self-worth, it is important to know how much value a child places on competence in certain domains in order to assess the extent to which particular areas of competence inform the child's sense of self-worth. Similarly in assessing the child's evaluations of acceptance by others, it is important to know what value the child places on particular relationships. Thus, if scales of perceived competence are used they should ideally be complemented by interviews that address these issues.

Administrative issues. Most of the recommendations regarding the instruments used in the NLSY79 have been for more complete assessment instruments. However, cost, respondent burden, and ease of administration also need to be considered in selecting instruments for ECLS. More complete assessments result in higher cost and respondent burden and can complicate the interviewer's task. Perhaps one solution would be to use more complete assessments with a representative subsample of children, while using more abbreviated versions with the larger sample. Oversampling of special populations would ensure sufficient numbers for group comparisons.

Timing is also an issue and needs to be taken into account in designing the study. Given the range of assessments being considered, it may be desirable to schedule assessments at different times. The length of particular assessments needs to be tailored to the age of the children being studied. Children may become tired and their attention may wander; consequently, assessments need to be relatively short in length.

References

Achenbach, T.M. & Edelbrock, C.S. (1981). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin 85*, 1275-1301.

Achenbach, T.M. & Edelbrock, C.S. (1983). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile*. Burlington, Vermont: University of Vermont.

Alexander, K.L. & Entwisle, D.R. (1988). Achievement in the first two years of school: Patterns and processes. *Monographs of the Society for Research in Child Development, 53*, Serial No. 218.

Baker, P.C., Keck, C.K., Mott, F.L., & Quinlan, S.V. (1993). *NLSY child handbook: A guide to the National Longitudinal Survey of Youth 1986-1990 child data.* Columbus, OH: Center for Human Resource Research, The Ohio State University.

Baydar, N. & Brooks-Gunn, J. (1991). Effects of maternal employment and child-care arrangements on preschoolers' cognitive and behavioral outcomes: Evidence from the children of NLSY. *Developmental Psychology, 27*, 932-945.

Bayley, N. (1969). *Bayley Scales of Infant Development: Birth to Two Years*. New York: The Psychological Corporation.

Bierer, B. (1981). *Preference for challenge among children who over-rate, under-rate, and accurately rate their cognitive competence*. Unpublished Ph.D. Dissertation, University of Denver.

Caldwell, B.M. & Bradley, R.H. (1984). *Home observation for measurement of the environment.* Little Rock, Arkansas.

Chase-Lansdale, P.L., Mott, F.L., Brooks-Gunn, J. & Phillips, D.A. (1991). Children of the National Longitudinal Survey of Youth: A unique research opportunity. *Developmental Psychology, 27*, 918-931.

Entwisle, D.R. & Hayduk, L.A. (1982). *Early schooling*. Baltimore: Johns Hopkins University Press.

Gesell, A., & Armatruda, C. (1947). *Developmental Diagnosis*. New York: Harper and Row.

Graham, P.J. & Rutter, M.L. (1968). The reliability and validity of the psychiatric assessment of the child, II. Interview with the parent. *British Journal of Psychiatry, 114*, 581-592.

Harter, S. 1985. *Manual for the self-perception profile for children*. Denver, CO: University of Denver.

Harter, S. & Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development, 55*, 1969-1982.

Hess, R.D., Holloway, S.D., Dickson, W.P., & Price, G.G. (1984). Maternal variables as predictors of children's school readiness and later achievement in vocabulary and mathematics in sixth grade. *Child Development, 55*, 1902-1912.

Kagan, J. (1981). *The Second Year: The Emergence of Self Awareness*. Cambridge MA: Harvard University Press.

Kellam, S.K., Branch, J.D., Agrawal, K.C., & Ensminger, M.E. (1975). *Mental health and going to school: The Woodlawn program of assessment, early intervention, and evaluation*. Chicago: University of Chicago Press.

Menaghan, E.G. & Parcel, T.L. (1992). Social sources of stability and change in children's home environments: Effects of parental occupational experiences and family conditions. Paper presented at The RAND Conference on Economic and Demographic Aspects of Intergenerational Relations, Santa Monica, March 20-22.

Parcel, T.L. & Menaghan, E.G. (1989). Child home environment as a mediating construct between SES and child outcomes. Working Paper. Columbus, OH: Department of Sociology, Ohio State University.

Peterson, J.L. & Zill, N. (1986). Marital disruption, parent-child relationships, and behavioral problems in children. *Journal of Marriage and the Family, 48*, 295-307.

Rutter, M., Tizard, J. & Whitmore, K. (1970). *Education, health and behavior*. London: Longman.

Stevenson, H.W. & Newman, R.S. (1986). Long-term prediction of achievement and attitudes in mathematics and reading. *Child Development, 57*, 646-659.

Zill, N. (1988). Behavior, achievement, and health problems among children in step families: Findings from a national survey of child health. In E. M. Hetherington and J. Arasteh (eds.), *The impact of divorce, single parenting, and step-parenting on children*, pp. 325-368. Hillsdale, NJ: Erlbaum.

Zill, N. & Snyder, N. (1981). [Tabulations from public use data tape of 1981 Child Health Supplement to the National Health Interview Survey]. Unpublished raw data. Washington, D.C.: Child Trends.

2.3 Greensboro Early Schooling Study

Purpose of the Study

The Greensboro Early Schooling Study (Morrison, Griffith, & Williamson, 1993a, 1993b; Morrison, Griffith, Williamson, & Hardway, 1993) was initiated in the fall of 1990 with the first of three successive waves of kindergarten children in Greensboro, North Carolina. The study was designed to examine characteristics of children, families, and schools that predict school readiness and early school success. The study assessed children's family and classroom environments and their performance on vocabulary, reading, general knowledge, and mathematics tasks. The cognitive assessments were administered in the fall and spring of the kindergarten year and in the spring of first and second grade for the oldest children; the youngest children were tested through the spring of first grade.

A related study, begun in the fall of 1991, examined the effects of extended-year schooling on children's cognitive growth (Frazier & Morrison, 1994). The study was conducted with two cohorts of kindergarten children at both traditional and extended-year magnet schools and used a similar set of assessment instruments. The cognitive assessments for the extended-year study were administered in the fall and spring of the kindergarten year and in the fall and spring of first grade. The first-grade fall assessments were introduced so that cognitive gains or declines over the summer months could be compared for the two groups. The extended-year students, who had attended school for 30 extra days during the summer, were expected to perform better on the cognitive assessments than traditional students who had not attended school during the summer months.

Both studies are small in size, with samples drawn from a particular city and school district. Their importance for ECLS lies not in the size and representiveness of the samples, but rather in the issues addressed and the assessment instruments used.

Sample Design

The sample for the Greensboro Early Schooling Study consists of 540 kindergarten children from three successive cohorts. The oldest children were followed through the spring of second grade; the youngest children were followed through the spring of first grade. All students are enrolled in elementary schools in the Guilford County Public School District in Greensboro, North Carolina. The sample includes approximately equal numbers of white and black students, is also balanced with respect to students' gender, and includes students from a variety of socioeconomic backgrounds.

The extended-year study was also conducted in Greensboro. Two cohorts of kindergartners from four traditional (180 day) magnet schools ($\underline{N}$ = 90; Cohort 1 $\underline{N}$ = 31; Cohort 2 $\underline{N}$ = 59); and one extended-year (210 day) magnet school ($\underline{N}$ = 91; Cohort 1 $\underline{N}$ = 34; Cohort 2 $\underline{N}$ = 57) were selected for study. The four traditional schools each had a different instructional emphasis: acceleration and enrichment, communications, open education, and science and

technology. The extended-year program offered a global studies magnet. A matched control group was used in an effort to sort out the effects of each potentially confounding component of the extended-year program. The composition of the traditional and extended-year groups were equivalent on 18 variables: child's IQ, school entrance age, gender, race, preschool experience, home literacy environment (derived from a composite of 13 variables), child's health, medical problems, birth complications, resident guardians, parents' education, parents' occupational status, parents' age, father's employment status, and parents' expectations for children's schooling. Students in the matched group were also participating in the larger Greensboro study. The matched group of traditional students for Cohort 1 was selected from an initial group of 34 students who returned parental consent forms; the matched group of traditional students for Cohort 2 was selected from an initial group of 71 students who returned parental consent forms. Students in Cohort 1 began kindergarten in the fall of 1991; students in Cohort 2 began kindergartenin the fall of 1992. As noted, all schools included in the extended-year study were magnet schools. District guidelines for magnet school attendance call for admission on a first-come, first-served basis; exceptions are made when particular admissions enhance the racial balance of school populations. As a result of this policy, magnet school students are representative of the school district's general population.

Assessment Instruments and Procedures

In both the Greensboro Early Schooling Study and the extended-year study, information was collected on characteristics of children, families, and schools thought to predict cognitive outcomes and school success. A set of cognitive assessments was also administered to children to measure their performance on vocabulary, reading, general knowledge, and mathematics tasks. The instruments used in the two studies are described below.

Parenting questionnaire. A background questionnaire designed to obtain information on parental education and occupation, family composition and structure, child's health, child's preschool experience, and family literacy environment was completed by parents. Questions regarding the quality of the home literacy environment included the number of child and adult magazines subscribed to; the number of newspaper subscriptions; ownership of a radio, television, stereo, dictionary and/or set of encyclopedias; hours of television watched per week by the child; possession of a library card by a member of the household and frequency of its use; and the number of people who read to the child, and how often.

A more comprehensive parenting questionnaire has been developed by Morrison and his students for use in future studies. The questionnaire addresses five dimensions of parenting that have been highlighted in research exploring how family influences children's academic development: knowledge and beliefs; literacy environment; rules, standards, and limits; family organization; and the affective climate of the home. Items frequently tap more than one dimension. For example, questions regarding parents' attitudes and beliefs about education may also be useful in assessing the quality of the home literacy environment and the affective climate. The expanded questionnaire has been piloted with approximately 100 middle- and upper-middle-class families with good results. Preliminary examination of response patterns revealed substantial

variability on several items, suggesting that the questionnaire will yield vertical answers from respondents. However, more extensive piloting is needed with a larger and more representative sample.

Several items examine parents' attitudes, beliefs, and knowledge about education and parenting. Parents are asked to evaluate the importance of listening to children, taking their opinions seriously, and encouraging children to explore, ask questions, and express their own opinions. Parents are also asked to assess factors that contribute to children's success in school (e.g., parental involvement, having a good teacher, amount of effort put into work, and innate ability), and to evaluate children's needs for discipline, guidance, and freedom to make mistakes and decisions. Other questions are designed to assess parents' attitudes toward children (e.g., respect for children, desire to spend time with them), their knowledge of developmentally appropriate behaviors (e.g., "I believe in toilet training a child as young as possible"), and their educational aspirations and expectations for their children.

Questions about the home literacy environment include the extent to which the parents read and enjoy reading, the extent to which the child reads or looks at books or has books read to him/her, the amount of television the child watches and the types of programs watched, the number of family trips to museums, plays, or concerts taken within the past six months, and the child's access to educational materials (e.g., books, magazines, games, puzzles) and resources (e.g., a computer, radio, television, and/or stereo).

Questions regarding family rules, standards, and limits focus on parents' typical responses to a variety of misbehaviors (e.g., disruptive behavior at school, hitting a playmate, lying), and on the consistency with which parents adhere to established rules and limits. Questions about family organization are designed to reveal the predictability and structure of home life and focus on family schedules and routines, including regular mealtimes and bedtimes, and established routines for each (e.g., setting the table, washing hands before mealtime, brushing teeth afterwards, taking a bath before going to bed). Parents are also asked to report on children's responsibilities around the house (e.g., household chores, yardwork, petcare), and on monetary incentives or rewards provided to children (e.g., money given as a reward for performing chores or withheld for failure to do required chores).

Questions addressing the affective climate of the family overlap with many of the items already mentioned (e.g., encouraging children's to explore, ask questions, express their opinions). Other items include parents' feelings of closeness to the child, appreciation shown to children for their accomplishments, help given in play or work activities, time spent alone with the child, and physical or vocal displays of affection or displeasure.

Teaching questionnaire. A teacher questionnaire was not used in the Greensboro study or the extended-year study. However, Morrison and his students have developed a teaching questionnaire that focuses on five dimensions of teaching that parallel the dimensions used in the expanded parenting questionnaire. Teachers are asked about their educational attitudes and beliefs (e.g., beliefs about the importance of various activities and teaching styles to children's academic

development—homework, study groups, drills, strict discipline); about the literacy environment of the classroom (e.g., frequency of class visits to the library, the number of field trips taken per year, seatwork assignments, displays of children's artwork); classroom rules, standards, and limits (e.g., typical reactions to student misbehaviors, consistency with which rules are adhered to, importance placed on good manners); classroom organization (e.g., classroom schedules and routines, including time spent on particular subjects, time spent in individualized or small-group instruction, and children's responsibilities for cleaning or straightening classroom areas); and the affective climate of the classroom (e.g., appreciation shown for students' accomplishments, interest in their problems, and respect for their feelings and opinions).

The Early Childhood Environment Rating Scale (ECERS). The ECER Scale (Harms & Clifford, 1980) was used to assess the developmental appropriateness of kindergarten classrooms. ECERS consists of six subscales: personal care routines, furnishings/displays for children, language-reasoning experiences, fine/gross motor activities, creative activities, and social development. Each subscale has a seven-point rating system (1 = inadequate; 7 = excellent); a rating of five indicates a "good" classroom environment. The classroom observations and ratings were made by researchers and were typically scheduled in the spring of the kindergarten year.

The Cooper-Farran Behavior Rating Scale. The Cooper-Farran Behavior Scale was completed by teachers in the fall and spring of kindergarten and the spring of first grade for all children in the Greensboro studies. This measure consists of two subscales: interpersonal social skills (e.g., questions regarding peer relations and peer interactions) and work-related social skills (e.g., questions regarding independence and cooperation).

The cognitive assessments and the self-concept scale. Several assessments were used to measure children's cognitive performance and their perceptions of their competence and social acceptance.

The Stanford-Binet Revised Intelligence Scale Short Form: The Stanford-Binet short form includes six subscales: vocabulary, bead memory, quantitative, sentence memory, pattern analysis, and comprehension.

The Peabody Picture Vocabulary Test-Revised (PPVT-R): The PPVT-R assesses children's receptive vocabulary; children are shown four pictures and are asked to identify the picture signified by the target word (e.g., cow or nest).

The PIAT-R General Information, Reading Recognition, and Mathematics Subscales: The General Information subscale of the revised Peabody Individual Achievement Test assesses the degree to which children have acquired culturally-relevant knowledge (e.g., "From what animal do we get milk?" "What are the colors of the American flag?"); the Reading Recognition subscale assesses children's skills in letter and word recognition; the Mathematics subscale assesses a wide range of math skills, including number recognition, addition, subtraction, multiplication, and division.

The Pictorial Scale of Perceived Competence and Social Acceptance for Young Children: The Pictorial Scale was developed for use with young children by Harter and Pike (1984). The overall scale includes four subscales that tap children's feelings of competence (cognitive and physical) and social acceptance (maternal and peer). Children are shown two pictures (e.g., a proficient speller and a child who is having difficulty spelling) and are asked to choose which child is most like themselves. Scores on each subscale range from one to four, with higher scores indicating a good self-concept.

Administration and timing of the assessments. The child assessments were individually administered in a quiet room on school grounds. In both the Greensboro Early Schooling Study and the extended-year study, the PPVTR, the PIAT, and the self-conceptscales were administered in the fall and spring of the kindergarten year and in the spring of first grade. For the extended-year study, the assessments were also administered in the fall of first grade so that the effects of the extended-schooling period could be evaluated. The Stanford-Binet Short Form was administered to all children during the middle of the kindergarten year.

Publications and Uses of the Data

Results from the Greensboro Early Schooling Study have been reported in three papers by Morrison and his colleagues (Morrison, Griffith, & Williamson, 1993a, 1993b; Morrison, Griffith, Williamson, & Hardway 1993). Results from the extended-year study are reported in Frazier and Morrison (1994). Several of the findings from these studies are presented below in the discussion of the assessment instruments.

A five-year follow-up study is planned for the 540 Greensboro children who participated in the original Early Schooling Study. A similar study with a new sample of children from School District 65 in Illinois (Evanston and a section of Skokie) has also been proposed. Approximately 400 kindergartners from two successive cohorts will be selected for study and will be followed for a five-year period. A follow-up study is also planned for the Greensboro extended-year study. A sample of 135 kindergartners (60 extended-year; 75 traditional) originally selected for study in the fall of 1992 will be followed over a five-year period. Since these studies are an extension of previous work, most of the assessment instruments used in the original Greensboro studies will also be used in the follow-up studies; the new parenting and teaching questionnaires (Morrison, Hardway, Frazier, & Stilson, n.d.) will also be used. Cognitive assessments will be administered to participating students every spring through the end of second grade. Once children reach third grade, their academic achievement will be measured by means of district-administered standardized tests (including measures on reading, mathematics, science, and social studies). For the Greensboro extended-year follow-up study, the cognitive assessments will be administered in both the fall and the spring of each year so that each ensuing period of summer instruction can be assessed. The children in the extended-year follow-up will continue to receive the cognitive battery through the fall of fourth grade since the sequencing to the district standardized tests does not permit the fine-grained assessment necessary to evaluate extended-year schooling. Data

collection for the Evanston study and the Greensboro follow-up studies is scheduled to begin in the fall of 1994.

Implications for ECLS

Several instruments used in the Greensboro studies seem appropriate for use in ECLS. Questions regarding the home literacy environment and the observational measures of classroom environments seem particularly promising. However, the cognitive assessments and the self-concept scale are problematic for reasons that will be noted below.

Parenting and teaching questionnaires. Because of their emphasis on the literacy environment of the home and the classroom, both the parenting and teaching questionnaires will be useful in developing questions for ECLS. While the parenting questionnaire includes items that are similar to those used on the HOME-Short Form (described in the study summary for the NLSY79 Child Assessments), it offers a more comprehensive assessment of the home literacy environment. Questions regarding the frequency with which parents read, parents' attitudes and beliefs about education, and their encouragement and supervision of their children's literacy activities tap dimensions of parent-child interaction that are missing in the HOME assessments.

As noted, the dimensions chosen for study have been highlighted in research exploring the influence of the family on children's academic development. Although factors such as low economic standing, minority group membership, and residence in a single-parentfamily have been linked with poor academic performance, research on other characteristics of family life suggests that a more comprehensive assessment of family structures and supports is needed to identify factors that contribute to children's academic success or failure. In a study of immigrant families, for example, Caplan, Choy, and Whitmore (1992) found that socioeconomically disadvantaged Indochinese children excelled academically in American schools. The researchers discovered that several elements of family structure (e.g., cooperative homework activities, predictable schedules, adherence to family rules) were critical in supporting children's academic progress.

Despite the comprehensiveness of the parenting and teaching questionnaires, both offer somewhat general assessments of family and teaching environments. Parents, for example, are presented with a series of hypothetical situations regarding children's problem behaviors (e.g., "You received a note from your child's teacher stating that your child has been disruptive at school. This is not the first time this has happened.") and are asked to indicate the likelihood of responding in particular ways to the behaviors described (e.g., let the situation go, take away a privilege, reason with the child). Parents are not asked to report on specific problem behaviors exhibited by their child or to indicate their responses to those behaviors. Similarly, parents are not asked to evaluate their child's academic performance or progress in school. They are instead asked to rate the importance of various factors to children's success in school (e.g., parental involvement, having a good teacher, amount of effort one puts into one's work; and the ability with which one is born). The teaching questionnaires are similarly structured. Teachers, however, are asked to complete a behavior rating scale for each child. In future studies,

information on children's attendance, grade-retentions, suspensions or expulsions, and referrals for behavioral or learning problems will also be obtained from school records.

Classroom observations. The Early Childhood Environment Rating Scale (ECERS) is a set of observational measures of the nature and quality of children's classroom interactions; both developmentally appropriate materials and activities are assessed. The six subscales—personal care routines, furnishings/displays for children, language-reasoning experiences, fine/gross motor activities, creative activities, and social development—provide a basis for comparing young children's classroom environments. As its name suggests, ECERS is intended for use in preschool and kindergarten classrooms; it would have to be modified for older children.

In comparing ECERS scores for traditional and extended-year classrooms, Frazier and Morrison (1994) found few differences in the quality of the classrooms observed. Although statistically significant differences between groups were found, most classrooms received relatively high overall scores. Other studies that have used the ECERS scale need to be identified so that its usefulness and reliability as a predictive measure can be evaluated.

A different classroom measure has been chosen for use in the Evanston study and the Greensboro extended-year follow-up study. For these studies, the classroom observations will be used to assess the classroom literacy environment directly. Observations will be conducted with the *Code for Instructional Structure and Student Academic Response* (Stanley & Greenwood, 1992), a 53-item observation system that allows recording of ecological and behavioral events within classroom settings. The CISSAR was designed primarily to address questions relating to classroom instruction and student academic behavior. Additional information on this coding system is needed to determine its appropriateness for ECLS.

The cognitive assessments. Although the cognitive assessments included in the Greensboro studies have been widely used, they have several limitations. The Revised Stanford-Binet Short Form was used to measure individual differences in skill levels early in the kindergarten year; the scores were then used to predict children's performance on other assessments. One problem with the Stanford-Binet and other intelligence tests is that they require a wide range of skills for the performance of a given task. In other words, skills within a particular domain may be obscured by the skills needed to comprehend the instructions for performing a task within that domain. Consequently, IQ tests are not as powerful as more domain-specific assessments in modeling individual growth curves. Second, results of the IQ tests appear to add nothing to the general pattern of findings that emerge from other assessments. In analyzing the results of the Greensboro Early Schooling Study, Morrison, Griffith, and Williamson (1993) found that IQ scores were strongly correlated with children's PPVT-R and Reading Recognition scores for the fall of the kindergarten year. In addition, they found the same pattern of broad individual differences in skill levels across the spectrum of predictor and outcome variables. A further problem with the Stanford-Binet is its length; like other intelligence tests, it is both time-consuming and costly to administer.

The findings from the Greensboro studies highlight a more general problem with the assessment instruments used. Morrison et al. (1993a, 1993b) found that not only were the PIAT and PPVTR scores highly correlated with IQ scores, they were also highly correlated with each other. For the fall kindergarten assessments, correlations ranged from .52 for receptive vocabulary and reading recognition scores to .78 for scores on the receptive vocabulary and general knowledge assessments. Morrison and his colleagues interpret these findings as indicative of general differences in individual skill levels; that is, children who score relatively low in one domain tend to score relatively low across all domains. However, correlations across domains may again reflect the instrument's failure to assess skills in only one domain. Children with poor verbal comprehension skills would score lower on all assessments that rely on complex verbal instructions.

Correlations between children's performance in the fall of kindergarten and the spring of first grade provide some support for this interpretation. Morrison et al. (1993a, 1993b) found that children's performance on the various assessments remained highly stable across the first two years of school, with one exception: fall kindergarten vocabulary scores were only modestly correlated ($r = .31$) with spring first grade reading scores. Since the Reading Recognition scale focuses on word and letter recognition rather than verbal comprehension, lower correlations between reading scores and the scores on the other assessments would be expected, particularly if there is an instructional emphasis on word and letter recognition during the early years of school. Stronger correlations would be expected among assessments that rely more heavily on verbal comprehension skills. The findings support these expectations.

Morrison and his colleagues note that not only do children enter kindergarten with widely differing skills, these differences are maintained and in some instances magnified across the first two years of school. On half the measures (reading and receptive vocabulary) individual differences were maintained; on the other half (cultural knowledge and mathematics) individual differences increased. Performance gains for children with higher initial skill levels also increased more sharply between the spring of kindergarten and the spring of first grade than those for children with poorer skills, suggesting that the two groups may differentially benefit from classroom instruction. One problem with this interpretation is that no assessments were done in the fall of the first grade. Consequently differential declines in performance across the summer months could not be measured. Research on seasonal learning effects suggests that socio-economically disadvantaged children benefit from instruction during the school year but show greater declines in performance across the summer months than less disadvantaged children (see, e.g., Alexander & Entwisle, 1988). In their own comparison of traditional and extended-year kindergartens, Frazier and Morrison (1994) found that extended-year students performed better than traditional students on fall first-grade math, reading, and general knowledge assessments. Previous research suggests that extended-year programs may be particularly beneficial to children who have few opportunities to develop skills outside the classroom. The magnification of differences in individual skill levels that were found between the spring of kindergarten and the spring of first grade may reflect differential opportunities to develop or maintain skills over the summer months rather than a failure of children with poorer skills to benefit from classroom instruction.

Currently the assessments for ECLS are planned for the fall and spring of the kindergarten year and for the spring of each succeeding year. If seasonal learning effects are to be addressed by ECLS, fall assessments should be considered for succeeding years of the study.

The Pictorial Scale of Perceived Competence and Social Acceptance for Young Children. The Pictorial Scale has been described in the study summary for the NLSY79 Child Assessments. Although the reliabilities for the overall scale are reasonably high, the reliabilities for the subscales are modest. In addition, the competence subscales (cognitive and physical) have been found to be less reliable than the social acceptance subscales (peer and maternal). Results from Frazier and Morrison's extended-year study (1994) raise further doubts about the reliability of these scales. Although extended-year students attended school for 30 additional days during the summer, self-reported levels of cognitive competence decreased over the summer for both traditional and extended-year students. Perceived competence decreased less for extended-year students but declines were still found to be statistically significant. According to the researchers, the school year for extended-year students began in late July and ended in mid-June. Given the relative shortness of the summer break, the reported declines in perceived competence for extended-year students are somewhat puzzling and raise further concerns about the reliability of the competence subscales.

References

Alexander, K.L. & Entwisle, D.R. (1988). Achievement in the first two years of school: Patterns and processes. *Monographs of the Society for Research in Child Development, 53*, Serial No. 218.

Caplan, N., Choy, M.H., & Whitmore, J.K. (1992). Indochinese refugee families and academic achievement. *Scientific American, February*, 36-42.

Frazier, J.A. & Morrison, F.J. (1994). Extended-year schooling and growth of language, academic, and general knowledge skills. Manuscript submitted for publication.

Harms, T. & Clifford, R. (1980). *Early childhood environment rating scale*. New York: Teachers College Press.

Harter, S. & Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development, 55*, 1969-1982.

Morrison, F.J., Griffith E.M., & Williamson, G.L. (1993a). Two strikes from the start: Individual differences in early literacy. Paper presented at the biennial meeting of the Society for Research in Child Development, New Orleans, March.

Morrison, F.J., Griffith E.M., & Williamson, G.L. (1993b). Two strikes from the start: Individual differences in early literacy skills. Paper presented at the second annual Head Start Research Conference, Washington, D.C., November.

Morrison, F.J., Griffith E.M., Williamson, G.L., & Hardway, C.L. (1993). Family literacy environment, learning-related social skills, and academic achievement. Paper presented at the second annual Head Start Research Conference, Washington, D.C., November.

Morrison, F.J., Hardway C.L., Frazier, J.A., & Stilson, S.R. (n.d.) Parenting and teaching questionnaires. Unpublished manuscript, Loyola University of Chicago.

Stanley, S.O. & Greenwood, C.O. (1992). *Code for instructional structure and student academic response (CISSAR): Observation manual*. (2nd ed). Kansas City, KS: Juniper Gardens Children's Project.

2.4 Prospects: The Congressionally Mandated Study of Educational Growth and Opportunity

Purpose of the Study

As part of the Hawkins-Stafford Amendments to the Elementary and Secondary Education Improvement Acts, Congress mandated the U.S. Department of Education (ED) to conduct a longitudinal evaluation of the short and long-term effects of significant participation in Chapter 1 (now called by the original name, Title 1) programs on student outcomes. The mandate specified that the evaluation should yield national estimates of student outcomes that can also be reported for the four primary census regions (Northeast, South, Midwest, and West) and three levels of urbanicity (urban, suburban, and rural). In an effort to capture both elementary and secondary grade spans, as specified by the mandate, the U.S. Department of Education adopted a three-cohort study design that provides student outcomes data for grades 1 through 12.

The three cohort design consists of first, third, and seventh grade student cohorts. Baseline data collection were conducted with the third and seventh grade cohort members during the spring of 1991. Baseline data were collected from the first grade cohort during the fall of 1991. The primary reason for scheduling the grade 1 cohort for fall data collection was to approximate a "true baseline measure" for students who are, for the most part, experiencing formalized learning processes for the first time. The collection of longitudinal data for each cohort is scheduled to be completed within a six-year period. The first follow-up for all cohorts was conducted in the spring of 1992 and subsequent follow-ups are scheduled for each spring through 1996.

The Prospects research design allows researchers to conduct time-series analyses of student outcomes using six data points over a five year period for each of the three cohorts. The design also includes some overlap in grade levels in which students are observed during the six year period. For example, during the course of the study, fifth grade data will be collected from both the first and third grade cohorts to permit researchers to compare changes in the implementation of Chapter 1 programs that may have occurred within a two-year period. During the design phase of the project, it was expected that changes resulting from 1993 Congressional reauthorization hearings would affect program implementation at a time where the cross-cohort grade overlap would be analytically useful. This assumption will be tested after the collection of the 1995 data.

Unlike other major studies conducted by the U.S. Department of Education (e.g., NLS-72, HS&B, and NELS) Prospects is primarily an evaluation study with the specific objective of providing an interim report[3] and final report (1997) to Congress assessing the impact of Chapter 1 on student achievement. Therefore, the sampling specifications were designed to support analysis objectives as opposed to providing a nationally representative multi-purpose data set that could be used to address numerous research issues.

The legislation states that the study should compare the educational achievement of educationally disadvantaged students with significant participation in Chapter 1 programs to

[3] The interim report was delivered in 1993.

students who did not receive Chapter 1 services. Thus, the Prospects sample was designed to include a representative sample of students who were likely to have sustained exposure to Chapter 1 services and a representative sample of comparable students who would not receive Chapter 1 services. Information gathered during the first two rounds of data collection seem to indicate that the concept of a naturally occurring "control group" of comparable students who do not receive Chapter 1 Services is somewhat problematic. Students who are not eligible for Chapter 1 services typically receive needed services through other programs sponsored by the school district.

Sample Design

A three stage, stratified sampling design was implemented for the Prospects study. Stratification was used to address the subgroup reporting requirements (i.e., estimates by region and level of urbanicity) and to increase sample efficiency at each stage. In the first stage, 120 districts were selected from all strata including the four census regions and the three levels of urbanization. Within strata, districts were drawn proportionate to a measure of size, which was based on the estimated number of economically disadvantaged students.

Upon selecting a sample of districts, district officials were contacted and asked to report the actual number of economically disadvantaged students and the number of limited-English-proficient (LEP) students within each of their schools. The operational definition for economically disadvantaged student given to district officials was "any student eligible for receiving free or reduced-price school lunch." Based on the information provided by the sampled districts, schools were stratified by their proportions of disadvantaged and LEP students. In the second stage of sampling, schools with higher proportions of poor and LEP students were drawn with higher probabilities.

In the third stage of sampling all students in the targeted grades within sampled schools were selected with certainty. Schools which had unusually large enrollments (N exceeding 120) in the targeted grades were subsampled. Elementary schools were subsampled in units of intact classrooms. In most instances, this meant selecting four classrooms from a list of five or more. At the middle or junior high schools, students in the seventh grade cohort were randomly selected from the entire roster of seventh grade students. A subsample of 75 students was selected for any middle school with an enrollment exceeding 120.

The Prospects sample design selected all students in the targeted grades and thereby designated no rules for excluding in-grade students from the baseline sample. However, a small portion of students were designated as ineligible for participating in classroom testing and questionnaire administration. Several weeks prior to the scheduled test and survey dates, school staff were given an opportunity to designate students who were LEP, learning disabled, or physically incapable of participating in classroom administrations; these students were only excluded from the test and questionnaire administration components of the study. LEP students who were proficient in Spanish, according to the judgment of school officials, were allowed to complete Spanish versions of the student questionnaire and also were administered a Spanish language achievement test battery (the Spanish Assessment of Basic Education or SABE).

BEST COPY AVAILABLE

The inclusion of other individuals in the study was determined by links to the sampled student. One parent of each sampled student was asked to complete a self-administered questionnaire. The regular classroom teacher of the sampled student was asked to complete a self-administered questionnaire on classroom instruction in math and reading/language arts/English. In addition, the Chapter 1 teacher/classroom aide (if one existed) reported on classroom instruction in the same subject areas. It is important to note that both parent and teacher data are intended to supplement the student data and should not be analyzed as separate representative samples. In this design, the student is the primary unit of analysis and the only other units to receive weights for baseline measures are the schools and districts selected in the initial sample.

In addition to parents and teachers, school principals were asked to complete two self-administered questionnaires and Chapter 1 District officials of sampled districts were also asked to complete a self-administered document. While Prospects collects longitudinal data from the student and other supplementary sources, responding teachers, principals, and district officials may change with each round of data collection.

The OBEMLA supplement. As specified in the Request for Proposal, the Office of Bilingual Education and Minority Language Affairs (OBEMLA) expressed interest in having separate analyses of LEP student data. In reviewing the core sample, sampling statisticians noted that the number of LEP students in the core sample was insufficient to support separate subgroup analyses. To increase the number of LEP students, OBEMLA funded a supplementary sample of 25 additional schools containing high concentrations of first and third grade LEP students. The LEP schools were incorporated into the core sample data collection plan and the full complement of data was collected for the LEP first and third grade cohorts. Based on recent information, OBEMLA expects to fund the supplement through the 1995 data collection period.

The Catholic school supplement. In the fall of 1991, The U.S. Catholic Conference elected to supplement the Prospects sample with a nonprobability purposive sample of 35 Catholic Schools. The supplement targeted the first and third grade cohorts and continued through the spring 1993 data collection. The timing of this supplement allowed the contractor to include the Catholic first grade cohort in the Prospects baseline data collection which was conducted in October and November of 1991. Baseline data for the Catholic school upper cohort was collected in the spring of 1992 and created a grade four cohort, which coincided with the first follow-up of Prospects' core third grade cohort.

Assessment Instruments and Procedures

The Prospects evaluation collected data from students and related sources in an initial nationally representative sample of 372 schools. The sample sizes for the first, third, and seventh grade cohorts were approximately 12,000, 12,000, and 7,000, respectively. All school based data collection was conducted over a three to four day period within one previously specified five day school week. In most cases, the first and third grade cohorts adhered to a four day testing session and the seventh graders were tested over a three day period.

The school based data collection sessions were scheduled according to the requirements for completing the various subtests of the achievement test battery. The test battery selected for this evaluation was the Comprehensive Test of Basic Skills Fourth Edition 1989 (CTBS/4). Using guidelines established by the test publisher, CTB Macmillan/McGraw-Hill, students were generally scheduled as intact classrooms for morning testing sessions, preferably Tuesday through Thursday or Friday. Mondays were avoided because of expected student absences and mornings were reserved because students are assumed to be fresher and less fatigued than during afternoons.

At least one afternoon during testing week was reserved to administer the Student Questionnaire. Three levels of questionnaires were developed for this assessment: a grade 3 through 5 questionnaire, a grade 6 through 8 questionnaire, and a grade 9 through 12 questionnaire. It was assumed during the design phase of the study that collecting questionnaire data in addition to test data would be too burdensome and somewhat difficult for the reading and comprehension levels of students in the first and second grades. It is important to note that during questionnaire administration to students in grades 3 and 4, contractor staff read the questionnaire aloud to keep students on tasks and to aid in the understanding of instructions, questions, and responses that may not have been easily understood by reading alone.

Other school based activities included distributing the self-administered documents: Parent Questionnaire, the Regular and Chapter 1 Teacher/Aide Questionnaire, the Student Profile Questionnaire, the Principal Questionnaire and the Characteristics of Schools and Programs Questionnaire. In addition to distributing these documents, contractor staff used afternoons to transfer student record information to the Student Record Abstract form.

Just prior to the start of the field period, district superintendents and district chapter officials received notification informing them of the start of the Prospects data collection period. The Chapter 1 District Coordinator (if one existed) or the Director of Research and Evaluation received the District Chapter Coordinator Questionnaire via express mail. This document was also self-administered with planned telephone prompting and in-person follow-up.

In general, all data collection activities either occurred within the school or emanated from the school (via distribution). The only exception is the district level document which was mailed directly to the district official from the contractor's office.

Trained contractor field staff conducted the school based data collection, while the Survey Administrators, assisted by Field Assistants, led the testing sessions. Based on guidelines developed by the test publishers, a Field Assistant was provided for every 12 students to field questions and monitor the pace of individual students. Student Questionnaire administrations were conducted under the same structure and Field Assistants were used to address questions regarding critical item edits.

The schedule for the collection of data was contracted to occur each spring beginning in 1991 and continuing through 1996 for each cohort, though 1996 data collection was eventually curtailed to permit more time and resources for preparation of the 1997 report to Congress. Just

prior to the 1994 data collection, ED decided to eliminate the seventh grade cohort from the study for two reasons. First, the seventh grade cohort was the least policy-relevant group, in that a small portion (less than 6 percent) of the students in the base year received Chapter 1 services and the percentage was expected to decline significantly in subsequent waves of data collection. Second, because of the wide dispersion of students in follow-up rounds, the cost of data collection increased; costs associated with collecting these data could not be justified given the limited relevance of this group to the Chapter 1 evaluation.

In summary, the final Prospects data will include five data points from the first and third grade cohorts. The seventh grade cohort data will include data spanning 1991 through 1994.

Domains assessed. Given that the primary foci of Chapter 1 compensatory services are math and reading/language arts/English, ED adopted a test instrument which provides comparable measures of student achievement in these areas across time and across cohort. Thus, the Comprehensive Test of Basic Skills Fourth Edition (CTBS/4) was selected to measure initial status and gains in the domains of Mathematics and Reading. The CTBS/4 is a vertically equated test and is organized by content areas that are routinely found in school curriculum guides throughout the nation.

In addition to the domains associated with standardized test measures, various data collection instruments were administered to various sources to obtain supplemental, descriptive and explanatory (predictive) information, as well as information concerning various outcomes. Listed below are several domains of interest by survey instrument.

District Level Questionnaire
Program Design, Management and Evaluation

School Program Questionnaire
Instructional Programs, Special Services, School Policies

School Principal Questionnaire
Staff Credentials, Administrative Leadership, Decision Making Techniques, Resource Allocation

Classroom Teacher Questionnaire
School Climate, Classroom Instruction, Coordination with Chapter 1 Instruction

Chapter 1 Teacher/Aide Questionnaire
School Climate, Classroom Instruction, Coordination with Regular Instruction

ESL/Bilingual Teacher Questionnaire

School Climate, Classroom Instruction, Coordination with Regular Instruction

Chapter 1 Counselor Questionnaire
Services Delivered, Counselor Background, Characteristics of Program

Student Questionnaire
Pre-School Experience, Course of Study, Grades and Performance, Activities, Family Background and Involvement, Opinions, Future Plans

Parent Questionnaire
Child's Demographics, Child at Home, Child at School, Parent's Contact with School, Family and Household Composition

Student Record Abstract
Locating Information, Special Programs, Disabilities, Services, Test Scores

Student Profile
Teacher's assessment of Student's ability, Student's Self-Image, and Report of School Related Behaviors and Salient Events

The instruments listed above were used with all three cohorts. It should be noted that no student questionnaire data were collected from the first grade cohort until these sample members reached the third grade. In addition, the baseline measure which occurred during the fall of 1991 included test measures only. The full complement of data was collected from relevant sources at the time of the first follow-up in the spring of 1992.

Special populations. Other than administering a Spanish translated student questionnaire and administering a Spanish language achievement test to students proficient in Spanish, no special arrangements or designations were made for students with special circumstances.

Publications and Uses of the Data

The two primary deliverables are the Interim Report and the Final Report to Congress. Three other descriptive reports are being prepared: Report on Chapter 1 Services, The LEP Student Report, and The Catholic School Report. The Interim Report (Puma, Jones, Rock & Fernandez, 1993) was the first major deliverable and compiled descriptive tables with breakdowns by poverty concentration and participation in compensatory education programs. Roberto Fernandez, sociologist at Northwestern University, wrote a small descriptive section on the limited-English-proficient (LEP) and Language Minority students (LM) for the OBEMLA Supplement to describe both student and school characteristics, school policies and practices, student performance measures for baseline and first follow-up, classroom practices, teacher qualifications, level of instruction, school climate, and parent involvement. While these are but a few domains of interest for the first report to Congress, it is important to stress that the report made no causal inferences between domains and student achievement. The causal modeling of student outcomes, as noted by ED, was reserved for the 1997 final report to Congress.

Implications for ECLS

Analysis conducted on the early waves of data demonstrates that the cognitive growth rates of advantaged students and less-advantaged students are similar during the academic year but vary greatly during the summer (Rock, 1994). Students from higher SES backgrounds continued to make cognitive gains during the summer months while those from lower SES backgrounds did not. During the school year, the lines plotting their rates of growth during the academic year were parallel—less advantaged students started at a lower level than other students but, during the academic year, progressed at the same rate as their peers. This finding -- consistent with earlier work by Heyns (1978) and Entwisle and Alexander (1988, 1992, 1994) -- suggests the ECLS should consider the possible utility of adding fall data collection points, at least for a subsample of students, and particularly at first grade, since measuring spring-to-fall achievement differences is critical both for the specific issue of summer learning differences and for the broader issue of school effects.

References

Alexander, K.L., & Entwisle, D.R. (1988). Achievement in the first two years of school: Patterns and processes. *Monographs of the Society for Research in Child Development, 53.*

Entwisle, D.R., & Alexander, K.L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review, 57*, 72-84.

Entwisle, D.R., & Alexander, K.L. (1994). Winter setback: The racial composition of schools and learning to read. *American Sociological Review, 59*, 446-460.

Heyns, B. (1978). *Summer Learning and the Effects of Schooling.* New York: Academic Press.

Rock, D.A. (1994). Results from Prospects. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April.

Puma, M.J., Jones, C.C., Rock, D.A., & Fernandez, R. (1993). *Prospects: The Congressionally mandated study of educational growth and opportunity. The interim report.* Washington, D.C.: U.S. Department of Education.

2.5 The D.C. Early Learning and Early Identification Longitudinal Study

Purpose of the Study

Beginning with the 1986-1987 school year, the District of Columbia Public School System (DCPS) initiated a three year study of its early learning programs in order to identify how such programs affect children's long-term school success. The study was prompted by high first grade retention rates among children attending D.C. public schools. Since the school system offered both preschool and kindergarten programs to children who were living in the District, the high retention rates for first graders prompted questions about the effectiveness of District pre-primary programs. The Early Learning and Early Identification Study was initiated both to identify the types of pre-primary programs that best prepare children for formal learning experiences and to identify the causes of learning deficits in primary grades so that preventive measures could be developed.

A follow-up study was conducted in 1990-1993 to investigate the effects of children's pre-primary experiences on school performance during the transition from primary to upper elementary grades. The study also attempted to identify predictors of grade retention and maladaptive behavior and to investigate how parental involvement and frequent moves affect children's academic achievement and adaptive behavior.

Sample Design

For the original study, three successive cohorts of children who attended preschool or Head Start programs in the D.C. Public School System were selected for study. Children who attended preschool/Head Start programs during the 1986-1987 school year were followed through the end of first grade; those who attended preschool/Head Start during the 1987-1988 school year were followed through the end of kindergarten. Children who attended pre-primary programs during the 1988-1989 school year were also studied but were not followed longitudinally.

A multi-stage sampling design was used for the initial selection of children. Preschool classrooms were first identified by region and classified by program type (e.g., child-initiated, academically focused, and intermediate). Children were then randomly selected from each of the three program types according to regional proportions of total pre-kindergarten and Head Start enrollments. Where too few classrooms in a particular region were available for study, additional children from other regions within the same model were randomly selected.

Complete information on sample size and composition is not available in the published reports we have received from the D.C. Public School System. However, it is clear from the demographic information provided in the published reports that the overwhelming majority of children in the D.C. school system are African American. In all samples, the number of African American children is near or exceeds 90 percent; the remaining children are predominantly Caucasian. Information on sample sizes for the original study was provided only for the third year (1988-1989) of the study. In that year, 286 preschoolers (Cohort 3) were selected for study.

However, Vineland forms (a measure of adaptive behavior) were returned for only 202 children (71%), and Progress Report forms (a measure of academic achievement) were returned for only 180 children (63%). The children for whom forms were completed came from 23 schools and 25 classrooms. Of the classrooms represented, 15 were child-initiated, 7 were intermediate, and 3 were academically-focused.

For the 1988-1989 kindergarten sample, 227 of 234 children who had been previously studied in preschool were found to be enrolled in the city's public kindergartens. Vineland forms were returned for only 113 of these children (50%), and Progress Report forms were returned for only 111 (49%). In addition, data were collected on 49 children who had not attended preschool the previous year. These children were matched by sex, ethnicity (when possible), and kindergarten teacher with Cohort 2 children to determine what effect preschool attendance had on performance in kindergarten. The kindergarten children for whom forms were completed came from 26 schools and 29 classrooms. Of the 29 kindergartenclassrooms represented, 9 emphasized socioemotional development and 20 emphasized academic preparation. Information on the composition of the sample by type of preschool program previously attended was not provided. A greater percentage of the follow-up sample came from single-parent families than did the cohort's original sample. Otherwise age, ethnicity, and pattern of school attendance were similar for Cohort 2 samples across both years of the study.

For the first grade sample, 234 of the 285 children who had been studied as kindergartners were found to be enrolled in DCPS first grade classes at the beginning of the 1988-1989 school year. A total of 186 Vineland forms (56%) and 264 Progress Report forms (79%) were returned by teachers for this group. Vineland data were also collected for 68 children who had no pre-kindergarten experience; Progress Report forms were collected for 96 children who had no pre-kindergarten experience. These children had been matched with Cohort 1 children during their kindergarten year (1987-1988) on the basis of sex, ethnicity (when possible), and kindergarten teacher. The first grade children for whom forms were completed came from 60 schools and 101 classrooms. Approximately half the children had attended kindergartens with a socioemotional emphasis; the other half had attended kindergartens that emphasized academic preparation. Information on the composition of the sample by type of preschool program previously attended was not provided. The first grade follow-up sample was more economically disadvantaged than the original Cohort 1 sample and was less ethnically diverse. Such differences were expected because more affluent or upwardly mobile families often withdrew their children from the public school system after the completion of pre-primary programs. In analyzing data for this cohort, economic differences in sample composition were controlled for by using eligibility for subsidized lunch programs as a covariate.

For the 1990-1993 follow-up study, data were collected on 461 children. Of these children, 81 percent ($n = 372$) had previously attended pre-kindergarten or Head Start programs in the D.C. Public School System. The remaining 89 children had first entered school as kindergartners. At the time of the follow-up study, the children were enrolled in 95 different elementary or middle schools in the district. Of these 461 children, 60 percent were originally from Cohort 1 (children who had attended preschool/Head Start programs in 1986-1987 and/or

kindergarten programs in 1987-1988) and 40 percent were from Cohort 2 (children who had attended preschool/Head Start programs in 1987-1988). No information on the composition of the sample by type of preschool and/or kindergarten program previously attended was provided in the report on the follow-up study.

The following data were collected for children from Cohort 1:

(1) 'Year 5' grades (n = 164 pre-K children; n = 71 K-only children) and achievement test scores (n = 132 pre-K children; n = 50 K-only children);

(2) 'Year 6' grades (n = 184 pre-K children; n = 89 K-only children);

(3) 'Year 7' Vineland adaptive behavior scores (n = 146 pre-K children; n = 66 K-only children).

For children from Cohort 2, the following information was collected:

(1) 'Year 5' grades (n = 177 pre-K children) and standardized achievement scores (n = 139 pre-K children)

(2) 'Year 6' Vineland adaptive behavior scores (n = 149 pre-K children).

If no grade retentions had occurred, 'Year 5' corresponded to third grade, 'Year 6' to fourth grade, and 'Year 7' to fifth grade.

Assessment Instruments and Procedures

For both the original and follow-up studies several measures were used to assess children's socioemotional development and academic performance. A survey of teachers' educational beliefs and practices was used to identify types of pre-primary programs attended by children in the study. The instruments used in both studies are described below.

Background information. For the original study, information was obtained on children's age, gender, ethnicity, absences from school, eligibility for a subsidized lunch program, and family status (single-parent vs. two-parent family). For the 1990-1993 follow-up study, information was also obtained on special education services received by children; previous grade retentions; transiency, as measured by moves from one school to another during a child's school career; and extent of parent involvement in children's school experience. Parent involvement was measured through teacher reports of parent-teacher conferences, home visits, parent visits to the classroom, and parent assistance in class activities.

Teacher survey of beliefs and practices. The teacher survey was used with preschool/Head Start, kindergarten, and first grade teachers in the original study to assess teachers' beliefs and practices with respect to early childhood education. The survey included

seven items addressing teachers' beliefs and seven items addressing classroom practices. Each item consisted of a 10-point rating scale measuring either the strength of teachers' beliefs about the appropriateness of particular instructional practices (e.g., direct instruction versus active learning experiences) or the extent to which teachers implemented such practices in their classrooms. Factor analysis of responses to this survey identified three types of preschool and kindergarten classrooms: programs emphasizing socioemotional development and child-initiated learning, programs emphasizing academic preparation, and programs falling between these two extremes. Two types of kindergarten programs were identified using similar procedures: moderately academic programs with a socioemotional emphasis and moderately academic programs with an emphasis on preparation for formal learning experiences.

Teacher interviews. In the original study, children's pre-K/Head Start, kindergarten, and first grade teachers were interviewed to determine the extent of contact they had with each child's parent(s) during the school year. Categories of contact included parent-teacher conferences, home visits by the teacher, extended class visits by the parent, and parental help with class activities. In the follow-up study, 'Year 5' teachers of Cohort 1 students and 'Year 6' teachers of Cohort 2 children were also interviewed. At each grade level, two groups of children were identified based on low (0 or 1 category fulfilled) or high (3 or 4 categories fulfilled) parent-school contact. Indicators of school competence, academic achievement, and children's development were analyzed for effects of parent involvement.

The Vineland Adaptive Behavior Scales (American Guidance Service). Completed by teachers for children in the study, the Vineland scales were used in both the original and follow-up studies to measure children's socioemotional development. The scales yield an overall Adaptive Behavior Composite Score as well as three domain scores measuring Communication Skills (receptive, expressive, written), Daily Living Skills (personal, domestic, community), and Socialization (interpersonal relationships, play and leisure time, coping skills). A fourth optional Vineland domain, Maladaptive Behavior, was used only in the 1990-1993 follow-up study.

The Metropolitan Reading Test (MRT). The MRT was used with kindergartners in the original study (1987-1988), when District policy required that the test be administered at the beginning of the kindergarten year to assess readiness for formal education. The MRT yields an overall composite score of reading readiness, as well as three domain scores measuring auditory, visual, and language components of reading readiness. No standardized assessment of math readiness was made.

The DCPS Early Childhood Progress Report. DCPS Early Childhood Progress Report forms were completed at the end of the school year by preschool and kindergarten teachers for children in the original study. Measures of children's classroom performance were based on DCPS criteria for mastery of basic skills. Progress report ratings were converted to a numerical grade point average with subscores measuring math/science, verbal (reading preparation, listening and speaking, literature), social (work and social habits), and physical skills.

The DCPS Report of Pupil Progress for Elementary Grades 1A-6B. This progress report form was used for the first grade sample in the original study and for all children in the 1990-1993 follow-up study. The form was completed by teachers at the end of the school year and was used to monitor children's mastery of basic skills. The sub-areas covered by the form include math, reading, language, spelling, handwriting, social studies, science, art, music, health/ physical education, and citizenship.

DCPS Competency Based Curriculum Objectives checklists. These checklists were used as an additional measure of children's progress toward mastering basic skills in reading and mathematics. The checklists were completed by first-grade teachers for children in the original study and by third and fourth grade teachers for children in the 1990-1993 follow-up study. The objectives vary by semester and year in school. The second semester fourth grade checklist includes 16 reading objectives and 33 math objectives. Fourth grade reading objectives include identifying phonetically irregular words, distinguishing denotative and connotative meanings, and constructing a topic outline. Fourth grade math objectives include plotting points on a grid, adding and subtracting whole and mixed numbers, and converting measurements (e.g., from feet to inches).

The Comprehensive Test of Basic Skills (CTBS). McGraw-Hill's Comprehensive Test of Basic Skills was used in the 1990-1993 follow-up study as a standardized assessment of school achievement. The CTBS is administered to all third-grade children in the D.C. Public School System. In addition to a Total Battery score, achievement is measured in the areas of reading (word attack, vocabulary, comprehension), language (spelling, language mechanics, language expression), mathematics (math computation, math concepts, and application), science, and social studies.

Publications and Uses of the Data

The final report of the original three-year study (Marcon, 1990) and the report for the 1990-1993 follow-up study (Marcon, 1994) are available from the Center for Systematic Educational Change, Early Learning Years Branch of the D.C. Public School System. Reports for the first two years of the original study were also prepared and can be obtained from the same source. Results from both studies have been used by educational administrators within the District of Columbia to evaluate early childhood education programs and to recommend reforms in D.C. public schools. Recommended reforms include eliminating academically-oriented pre-primary programs in favor of programs emphasizing socioemotional development; introducing continuous progress/ungraded primary programs as an alternative to retaining children in grade; and encouraging greater parental involvement in children's schooling by using strategies that have been effective with Head Start parents.

Implications for ECLS

In focusing on the effectiveness of various models of early childhood education, the D.C. early learning studies raise a number of important questions regarding both the developmental appropriateness of preschool and kindergarten curricula and children's mastery of curriculum objectives. Specific questions are addressed below.

Developmentally appropriate classrooms. The researchers for both the original and follow-up studies conclude that overly academic early childhood programs are developmentally inappropriate and have a negative impact on children's later academic achievement and social development. They note the following.

> By fourth grade, children who had attended academically-directed Pre-K programs were earning noticeably lower grades and passing fewer fourth grade reading and mathematics objectives, despite adequate performance on third grade standardized achievement tests. By fourth and fifth grades, children from academic Pre-K programs were developmentally behind peers and displayed noticeably higher levels of maladaptive behaviors. (Marcon, 1994: 63).

Although academically-directed preschool programs appear have a negative impact on later achievement and behavioral outcomes, the specific effects of such programs remain unclear. Since no baseline measures were used to measure individual differences in adaptive behavior and cognitive development at the beginning of the preschool/Head Start year, it is difficult to determine whether the findings reflect differences that existed prior to preschool or that result from differences in instructional practices. If instructional practices do contribute to later academic and behavioral outcomes, what specific mechanisms are involved? Do differences in instruction alone account for different outcomes, or do other factors contribute to this pattern of differences? In examining the relationship between the type of preschool program attended and the extent of parents' involvement in children's schooling, researchers found that parents were less likely to be involved in school activities if their children attended academically-directed preschool programs; the researchers suggest that such programs may discourage parents' involvement. Less involvement by parents would in turn have a negative impact on children's school success. However, it remains unclear how, and to what extent, such programs discourage involvement by parents.

The studies also raise questions about children's educational experiences between preschool and fourth grade. If children's preschool experiences continue to influence academic achievement and behavior, are children's initial school experiences reinforced by their experiences in later grades? If so, in what ways? Do children who perform poorly in earlier grades have lower performance expectations as they progress in school? Do these children receive support for academic activities outside the classroom? Do schools offer special services to children who are

having academic or behavioral difficulties? Do such programs unintentionally label children and place them at a further disadvantage academically?

In attempting to trace the possible links between early school experiences and later academic and behavioral outcomes, it becomes clear that several factors need to be taken into account. The relationships among these factors also need to be explicitly modeled (Alexander & Entwisle, 1988). The identification of factors that potentially contribute to school success and the conceptualization of their interrelationships should ideally guide the design of studies that attempt to evaluate the effectiveness of particular programs. Conclusions regarding the developmental appropriateness of particular programs should be based on measures that are developed specifically for that purpose, rather than being inferred from student outcomes.

Curriculum-based checklists. In the D.C. studies, competency-based checklists were used to measure students' progress toward mastering particular curriculum objectives in the areas of reading and mathematics. These checklists may prove useful in developing curriculum-sensitive measures for ECLS. Such measures often assess academic achievement more accurately than standardized tests, particularly for students who respond poorly to testing conditions. Such students may have mastered particular skills but fail to demonstrate that mastery on standardized tests.

Copies of the curriculum-based checklists used in the D.C. studies were not available for review. Examples provided in the reports of the original and follow-up studies suggest that the items used to assess particular competencies may be too specific for ECLS, because items would need to be general enough to use with a wide range of school systems. Ideally the items would emphasize mastery of concepts and general skills and principles rather than isolated facts or specific rules. The actual checklists should be reviewed with these criteria in mind so that the appropriateness of particular items can be assessed.

References

Alexander, K.L. & Entwisle, D.R. (1988). Achievement in the first two years of school: Patterns and processes. *Monographs of the Society for Research in Child Development, 53*, Serial No. 218.

Marcon, R.A. (1990). *Early learning and early identification: Final report of the three year longitudinal study*. Washington, D.C.: District of Columbia Public Schools.

Marcon, R.A. (1994). *Early learning and early identification follow-up study: Transition from the early to the later childhood grades 1990-1993*. Washington, D.C.: District of Columbia Public Schools.

2.6 National Education Longitudinal Survey, 1988 (NELS:88)

Purpose of the Study

Beginning in 1988 with a cohort of 26,432 eighth graders attending 1,052 public and private schools across the nation, NELS:88 was designed to provide longitudinal data about critical transitions experienced by students as they leave eighth grade school settings, progress through high school (or drop out), enter and leave postsecondary institutions, and enter the work force. The 1988 eighth-grade cohort has been followed at two-year intervals (specifically, first follow-up—1990; second follow-up—1992; third follow-up--1994), and a fourth follow-up tentatively scheduled to take place in 2000. Major features of NELS:88 include:

- the integration of student, dropout, parent, teacher, school administrator, and school records (transcript) surveys (NELS:88 components are depicted in table 6, below);
- curriculum-sensitive cognitive tests in reading, mathematics, science, and social studies;
- the inclusion of supplementary components to support analyses of educationally or demographically distinct subgroups (for example, oversamples of Asians and Hispanics, as well as students in private schools); and
- the design linkages to previous longitudinal studies (High School and Beyond [HS&B], the National Longitudinal Study of the High School Class of 1972 [NLS-72]) and other current studies (for example, the National Assessment of Educational Progress [NAEP] testing program and high school transcript data collections).

Sample Design

NELS:88 employed a two-stage base year sample design. In the first stage, stratified disproportionate samples of schools were selected from frames consisting of public and private schools in the 50 states and the District of Columbia that contained eighth-grade students. In the second stage, random samples were selected from frames of eighth graders, with oversampling of Hispanic and Asian eighth graders. (For details, see Spencer, Frankel, Ingels, Rasinski & Tourangeau, 1990). Some subsampling took place between the 1988 base year and the 1990 first follow-up, when eighth graders had dispersed to numerous high schools. The 1990 and 1992 samples were freshened to render them fully representative of the nation's 1990 sophomores and 1992 seniors.

Table 6: Base year through fourth follow-up—NELS:88 components

	Base Year	First Follow-up	Second Follow-up	Third Follow-up	Fourth Follow-up
Data collection	Spring term, 1988	Spring term, 1990	Spring term, 1992	Spring, 1994	Spring, 2000
Grades included	Grade 8	Modal grade Sophomore	Modal grade Senior	H.S. + 2 years	H.S. + 8 years
Cohort	Students: questionnaire, tests	Students, dropouts: questionnaire, tests	Students, dropouts: questionnaire, tests, transcripts	All individuals: interviews	All individuals: interviews, post-secondary transcripts
Parents	Questionnaire	None	Interview	None	None
Principals	Questionnaire	Questionnaire	Questionnaire	None	None
Teachers	Two teachers per student (English, social studies, math or science)	Two teachers per student (English, social studies, math or science)	One teacher per student (math or science)	None	None

There were rare exclusions from the base year school sampling frame, namely, Bureau of Indian Affairs schools, special education schools for the handicapped, area vocational schools not enrolling students directly, and schools for dependents of U.S. personnel overseas.

At the student level, implicitly, students in ungraded programs were excluded, since grade rather than age was used to define the cohort of interest. Schools were allowed to exclude from the sample students who had mental, physical, or linguistic barriers to participation. Some 5.4 percent of the potential sample was excluded. A special followback study re-examined the enrollment and eligibility status of base year (1988) excluded eighth graders two and four years later (Ingels 1991, 1996).

Assessment Instruments and Procedures

This summary emphasizes the NELS:88 cognitive test battery (Rock & Pollack, 1991, 1995; Ingels, Scott, Rock, Pollack, & Rasinski, 1994). Student questionnaire data, teacher data and teacher ratings of students, parent reports, academic transcripts, and school administrator data were also collected for multiple time points. Some ecological variables, for example community characteristics taken from Census files, also appear on the data set. In addition to providing information about such possible correlates of achievement as home background, parental

BEST COPY AVAILABLE

involvement, student self-esteem, classroom behavior, and so on, in some cases, there is direct articulation between questionnaire items and test content (for example, the teacher questionnaire collects specific information about the sampled student's classroom including topics taught within subject, so that opportunity to learn can be related to test performance).

The NELS:88 cognitive test battery spanned three grades (eighth, tenth, and twelfth) in four content areas: Reading, Mathematics, Science, and Social Studies (History/Citizenship/ Geography). The tenth and twelfth grade mathematics and reading tests incorporated multi-level forms differing in difficulty. In tenth grade, eighth grade reading and mathematics test results were used to assign students to a form of appropriate difficulty. A like procedure was repeated in the twelfth grade. The tenth and twelfth grade science and social studies tests were grade-level adaptive in the sense that everyone took the same form within a grade but each succeeding grade level form included additional more difficult items.

Objectives. The test specifications of the NELS:88 longitudinal test battery were dictated by its primary purpose: accurate measurement of the *status* of **individuals** at a given point in time, as well as their *growth* over time. Principal test objectives were as follows:

- Item selection should be curriculum-relevant, with emphasis on concepts, skills, and general principles. (When measuring change or developmental growth, overemphasis of isolated facts at the expense of conceptual and/or problem-solving skills may distort the gain scores due to forgetting.)

- The tests should not make the students feel pressured by time; the vast majority of students should be able to complete all the tests.

- There should be little evidence of floor or ceiling effects.

- Reliabilities of the component tests should be psychometrically acceptable for measuring individual status as well as growth.

- The accuracy of measurement (i.e., the standard error of measurement) should be relatively constant across SES, sex and racial/ethnic groups.

- The NELS:88 battery should be designed to reduce the gap in test reliabilities that is typically found between the majority group and racial/ethnic minority groups.

- The NELS:88 test battery should attempt to minimize Differential Item Functioning (DIF) across gender and racial/ethnic groups that arises from irrelevant content that favors one or more of the groups.

- The test content areas should demonstrate discriminant validity. That is, while the tests should be internally consistent and characterized by a large dominant factor,

they should yield a relatively "clean" although oblique four factor solution. The four factors should be defined by the four tested content areas.

- Subscores and/or proficiency scores should be provided where psychometrically justified. The tests were designed to provide behaviorally-anchored proficiency (mastery) scores in the areas of Reading, Mathematics, and Science.

- The NELS:88 test battery should share sufficient common items both across and within grade level forms, and with the HS&B battery, to provide articulation of scores for vertical equating in NELS:88 as well as cross-sectional equating with the 1980 HS&B sophomore cohort in mathematics.

- There should be sufficient item overlap between the National Assessment of Educational Progress (NAEP) mathematics test and the twelfth grade NELS:88 mathematics test to cross-walk to the NAEP mathematics scale.

- The reading test passages should provide relatively broad content coverage and have items that span at least three cognitive process areas. There also should be at least one passage that identifies in some way with minority concerns. Similarly, there should be at least one passage in which the main character is a female.

- The four content areas Reading, Mathematics, Science, and Social Studies (History/Citizenship/Geography) must be administered (including time for administration instructions) within 85 minutes.

- The tests should be sufficiently reliable to support change measurement, and be characterized by a sufficiently dominant underlying factor to support the Item Response Theory (IRT) model. This latter requirement is necessary to support the vertical (longitudinal) equating between retestings as well as (for math) the cross-sectional linking with HS&B and NAEP. IRT vertical equating puts scores within a given content area on the same scale regardless of the grade in which the score was obtained. This allows the user to interpret scores the same way whether they were from the eighth, tenth, or twelfth grade.

- Independent of the vertical scaling, the testing time constraints made achieving desired reliabilities problematic without introducing some sort of adaptive testing. In order to achieve this level of reliability, as well as reduce the possibility of floor and ceiling effects, the Mathematics and Reading tests were designed to be multi-level at the tenth grade and twelfth grade. Hence a further test objective was that there be sufficient linking items *across forms within grade* to allow equating using IRT models.

Test development process. The items that were used in the final eighth-grade forms were selected from a much larger pool composed of items from NAEP, HS&B, the Second International

Mathematics Study (SIMS), ETS test files from previous operational tests, and items specifically written for NELS:88. The selection of items for the field test item pool was based on the consensus of subject matter committees made up of curriculum experts.

The subject matter committees consisted of educators, teachers, and college professors specializing in middle school curricula. There was considerable personnel overlap with similar subject matter committees used in the NAEP item pool development. ETS test development specialists were in attendance and worked with their respective subject matter committees in developing the eighth, tenth, and the twelfth grade assessment objectives. Once the assessment objectives were agreed upon, the subject matter committee members classified the items according to the objectives. Fifty Reading items, 82 Mathematics items, 42 Science items, and 60 History/Citizenship/Geographyitems were selected for pretesting. Field tests were administered to eighth, tenth, and twelfth graders in the Spring of 1987 (Rock & Pollack, in Ingels et al. 1987). The results of the field testing were scrutinized by additional committees of subject matter experts who suggested numerous modifications in content, format, and wording of the items, and made judgments on content coverage. Final revisions and item selections were made by project staff on the basis of their input, and reviewed by NCES staff. Decisions about basic directions to follow in assessment design were made in meetings between contractor and NCES staff and reviewed by the Technical Review Panel.

The designs of the 1987, 1989, and 1991 field tests were as follows:

Base year (1987) field test:

- testing at grades eight, ten, and twelve, with a testing $\underline{N}$ = 1200 at each grade, with 600 examinees per form per grade.

- two test forms spiralled within grade:
  Form A = Reading (50 items); Science (42 items);
  Form B = Math (82 items); Social Studies (60 items)

First follow-up (1989) field test:

- A longitudinal sample of 200-300 Form A and 200-300 Form B eighth grade base year field test test-takers two years later,

- 200-300 Form A and 200-300 Form B tenth grade base year field test sample members two years later;

- and a freshening sample to provide a total of 400-500 observations per form per grade.

Second follow-up (1991) field test:

The longitudinal sample of 1987 eighth graders was augmented with 1991 seniors to provide for a sample of 2,070 test takers, distributed across five forms of the test. Additional test forms were employed so that free response items could be field tested as well as the multiple choice items that made up the longitudinal battery.

The base year field test item pool was over twice the size of the target final item pool (for example, 50 reading items were tested so that a test with 21 items could be selected.) Three grade levels were tested at once so that baseline items could be chosen that would have relatively high biserials (indicating scalability) and show reasonable longitudinal empirical gains (changes in difficulties), although items with high biserials that did not show significant gains remained candidates for inclusion if they had been identified as "marker" items for behavioral anchoring purposes.

While choice of the correct baseline items required use of data on linking items taken from three cross-sectional grade samples, ideally, item parameters will be re-estimated and refined as true longitudinal data become available. This approach, followed in the 1989 and 1991 NELS:88 field tests, provides item or task parameters that are optimal for measuring change. Also, note that while the field test employed matrix sampling, spiraling was used to maximize the item pool at a reasonable burden per student. The main study design did not use spiraling even though it is advantageous for taking a broader measurement of the curriculum and suitable for group level cross-sectional measurement (as in NAEP), because it would have been inoptimal for measuring individual change over time.

Test administration and special populations. The NELS:88 tests were group-administered by NORC field staff. Generally, accommodations were not made for special populations; a rare exception to this rule was made for students who could complete only a large-print version of the tests; for these students, the text of the tests would be enlarged in a copying machine. A further rare exception to this generalization was the case of emotionally disturbed students who could not be tested in a group setting; individual administration was made available in such instances. Sign language interpreters were not made available. No attempt was made to produce a Spanish-language, Braille, or other non-English version of the tests, although student and parent questionnaires were translated into Spanish (for further details of the methodology employed with Spanish-language instrumentation, see Ingels 1996). Sometimes students with disabilities were given extra time to complete a test. However, there are questions about comparability of results when extended testing time is offered for some students (Willingham et al. 1988). Since the NELS:88 field tests did not experiment with modes of accommodation and test their validity, special accommodations were avoided in the main study.

NELS:88 did, however, return to baseline excluded students in the 1990 and 1992 follow-up rounds in order to (a) reassess their eligibility (for example, a student who was nonproficient in English in 1988 might have become sufficiently proficient to complete survey instruments by 1992), (b) gather basic sociodemographic data about those excluded, so that potential

undercoverage biases could more fully be understood; and (c) to ascertain their enrollment status, so that national dropout statistics would reflect all members of the cohort, regardless of their ability to complete NELS:88 survey instruments. (A complete account of student eligibility and exclusion issues is provided in the NCES report *Sample Exclusion in NELS:88: Characteristics of Base Year Ineligible Students; Changes in Eligibility Status After Four Years* [Ingels, 1996].) In addition, high school transcripts were collected for some excluded students, a practice also adopted by NAEP in 1987 and 1990.

Presentation of scores. While the NELS:88 battery provides test scores with the usual normative interpretation (means, achievement quartiles, and so on), it was also designed to have "mastery" or "proficiency" level scores in mathematics, reading, and science. These multiple criterion-referenced levels serve two functions. First, they help with interpreting what a score level "means" in terms of what a child can or cannot do. Second, they are useful in measuring change at particular score points along the score scale. In particular, when certain school processes can be expected to be reflected in score changes taking place at specific points along the score scale, then changes in percent or probability of mastery at that point in the scale would be better measures of the impact of the school process on student growth than would changes in the overall test score. Three levels of proficiency were marked in the reading test, five in the mathematics test, and three in the science test, defined as follows:

Reading

Reading Level 1: Simple reading comprehension including reproduction of detail and/or the author's main thought.

Reading Level 2: Ability to make relatively simple inferences beyond the author's main thought and/or understand and evaluate relatively abstract concepts.

Reading Level 3: Ability to make complex inferences or evaluative judgments that require piecing together multiple sources of information from the passage.

Mathematics

Math Level 1: Simple arithmetical operations on whole numbers: essentially single step operations which rely on rote memory.

Math Level 2: Simple operations with decimals, fractions, powers, and roots.

Math Level 3: Simple problem solving, requiring the understanding of low level mathematical concepts.

Math Level 4: Understanding of intermediate level mathematical concepts and/or having the ability to formulate multi-step solutions to word problems.

Math Level 5: Proficiency in solving complex multi-step word problems and/or the ability to demonstrate knowledge of mathematics material found in advanced mathematics courses.

Science

Science Level 1: Understanding of everyday science concepts; common knowledge that can be acquired in everyday life.

Science Level 2: Understanding of fundamental science concepts upon which more complex science knowledge can be built.

Science Level 3: Understanding of relatively complex scientific concepts; typically requiring an additional problem-solving step.

Gain analysis can be conducted using IRT number-right scores, dichotomous proficiency scores, or continuous probability of proficiency scores.

Scaling and the construct validity of the NELS:88 content areas. While the multi-level adaptive approach used in mathematics and reading and the grade level adaptive approach used in the science and the social studies tests helped in minimizing floor and ceiling effects, it was decided that more recent developments in IRT models would be used to exploit the adaptive nature of the NELS:88 battery fully. More specifically, a Bayesian procedure (Muraki & Bock, 1991) estimated both the item parameters and the ability scores and allowed for separate prior ability distributions, which considered the differing ability distributions associated with the various forms used across and within grades.

While internal correlational analysis among the scale scores shows discriminant and convergent validity for the content areas, a further issue of critical importance is how well this Bayesian IRT approach (Muraki & Bock, 1991) worked compared to traditional techniques (LOGIST conditional maximum likelihood estimation). Validity for alternative approaches to IRT scaling as well as for the content areas themselves is defined here in terms of the pattern of correlations between IRT scores and relevant outside process and demographic variables. In the end, longitudinal studies that emphasize policy decisions must concern themselves with describing the extent of the relationship between student performance and school and home-based learning. Analysis of NELS:88 test data (Rock & Pollack, 1995) reveals that the normal prior Bayesian procedure showed stronger relationships between gains and virtually all the process/demographic variables than did the alternatives. As hoped and expected, NELS:88 aggregate (total) score gains showed expected patterns of overall gain while gains in proficiency probabilities show maximum relationships with school processes (for example, placement in a particular curriculum track) that targets learning that is appropriate for that particular mastery level.

Publications and Uses of the Data

Use of NELS:88 data—particularly longitudinal analysis—has just begun. As part of its NELS:88 third follow-up contract, NORC maintains a bibliography of NELS:88 publications. Already, as of June 1994, there were approximately 300 entries in the bibliography as of May 1996. However, limited *methodological* work has been published on the *test battery* beyond the base year and second follow-up psychometric reports and the first follow-up technical report. Two recent reports from the Stanford Center for Research on the Context of Secondary Teaching offer an alternative avenue to approaching the issues the NELS:88 proficiency scores are intended to address. Kupermintz, Ennis, Hamilton, Talbert & Snow (1994), in connection with the NELS:88 mathematics test, and Hamilton et al. (1994), in connection with the science test, stress that rather than using total scores alone, multidimensional achievement scores need to be used. NELS:88 mathematics achievement data support using subscores that yield differential relations with student, teacher, and school variables (for example, math knowledge and reasoning factors may be distinguished; student attitudes, instructional variables, course, and program experiences appear to relate more to knowledge, whereas sex, SES, and some ethnic differences seem to relate more to reasoning). They also illustrate the use of subscores reflecting the multidimensionality of the NELS:88 science tests, as well as the relationship of science subscores to student and teacher effects that total scores used alone would miss.

Implications for ECLS

Sample design. NELS:88 provides important lessons that address three critical ECLS sample design problems.

- NELS:88 offers a clear model for within-school oversampling of policy-relevant subgroups. (See Spencer, Frankel, Ingels, Rasinski & Tourangeau, 1990).

- NELS:88 supplies a clear means of dealing with the problem of school sample nonrepresentativeness in the follow-up rounds of a longitudinal survey. (See Spencer & Foran, 1991; Qian, 1996; Ingels, Scott & Frankel, 1996.)

- NELS:88 provides a workable means for "freshening" follow-up round samples to make the grade-level representative. (See Ingels, Scott, Rock, Pollack & Rasinski, 1994; Ingels and Owings, 1995).

Assessment. Several lessons can be drawn from the NELS:88 tests.

- Multilevel tests are often desirable to avoid floor and ceiling effects in longitudinal measurement. Using the same test form for students of different ability and achievement levels can seriously inflate the error of measurement; NELS:88 offers a useful model for tailoring test forms to a particular student's ability level. (See Rock & Pollack, 1995).

- The NELS:88 experience suggests that national longitudinal tests can benefit greatly from employing a multigrade baseline field test with a sample size sufficient to support matrix sampling and provide observations that reliably estimate item means, variance, and covariances as well as develop item response theory parameters and scales that can link performance across grades. (Note that part of the testing of the vertical scalability of items involves determining how well item traces fit *within* grades.) However, subsequent field tests should be used to supply *longitudinal* test data to refine task and item parameters that are optimal for change measurement. (See Rock & Pollack, in Ingels 1987).

- NELS:88 shows that if items can be developed to specifications that include criterion-referenced markers of different points in a generalized growth curve, then the researcher or policymaker can talk about changes over time in mastery or proficiency levels as well as normative change. These criterion-referenced points along the growth curve permit gain to be looked at qualitatively (that is, where on the scale change is taking place), not just in quantitative terms, and render both longitudinal and cross-sectional results more interpretable. (See Rock & Pollack, 1995).

- Little is known about the relationship between test validity and use of special accommodations for testing the handicapped, and there are many views of the desirability of full inclusion (Ysseldyke & Thurlow, 1993; Thurlow, Ysseldyke & Silverstein, 1993). Nevertheless, NELS:88 provides evidence (see Ingels 1996) for several important points: (1) eligibility can change over time—an important consideration in a longitudinal study that expects to freshen follow-up samples to make them grade-level representative; (2) there is much evidence that test inclusion and exclusion decisions on the part of school personnel often lack reliability or validity; (3) there are clearly means to obtain indirect information about individuals who cannot be directly assessed—information that may give evidence of important educational outcomes, and that, at the very least, provides a basis for estimating sample undercoverage biases and their impact on survey data.

- In regard to students with language barriers (NEP/LEP), NELS:88 was able to assess about half of the LEP population. Overall, about 1.5 percent of the potential eighth grade sample had to be excluded for language reasons. However, the followback study of excluded students showed that of those who were excluded for language reasons, the majority were capable of completing survey forms two or four years later (Ingels, 1996). This fact underlines the need to retain LEP/NEP students in longitudinal samples, even if they are unable to complete baseline tests. Of course, the number of NEP/LEP students is increasing and is highest at the lower grades. The 1992 NAEP identified 4 percent of the potential fourth grade sample as LEP and excluded from assessment 3 percent of the potential sample. At grade 8, 3 percent were identified as LEP, and 2 percent excluded. At grade 12, the 1992 NAEP identified 2 percent as LEP and excluded 1 percent (Mullis, Dossey, Owen & Phillips, 1993). CPS data for 1989 (*Condition of Education 1992*) show that of children 8 to 15 years old in school, 11.5 percent are language minority (speak a language other than English at home) and 3.2 percent are LEP (by family self-report). Using a different reporting source (state education agencies), the 1993 OBEMLA LEP

study (Henderson, Abbott & Strang, 1993) suggests that 5.6 percent of students nationwide are LEP (but 19 percent of students in California and New Mexico). Again, LEP proportions are always somewhat higher in the lower grades and proportions are growing over time. For a kindergarten study in 1998-1999, the NELS:88 strategy of allowing NEP and some LEP students to be excluded is not likely to be acceptable.

- Important features of NELS:88 that will benefit ECLS are teacher ratings of students and provision of classroom-level data, as well as multiple parent surveys. Community-level and other zipcode or tract-level census data have only recently been added to NELS:88 and have not yet been used in analysis, but it should be noted that addition of such ecological data was achieved at low cost with no burden to respondents.

- A probable criticism of the NELS:88 tests is that the longitudinal battery was limited to multiple choice items. This criticism should be expected despite the fact that trials and experiments in the 1991 field test and 1992 main study did allow free response items to be used and scored. Analysis of overall and subgroup results suggests that use of free response items in the longitudinal battery would not have provided a substantial body of additional discriminating information, although for selected subgroups this generalization may be less true. (See Pollack & Rock, forthcoming).

- NELS:88 provides a clear model of how to devise crosswalks between national testing programs (NELS:88 to HS&B, NELS:88 to NAEP) and suggests how valuable such cross-study equating can be.

- A major criticism of the HS&B tests was that they were more ability tests than truly curriculum-sensitive achievement batteries. NELS:88 tests should escape this criticism, given their thoughtfully elaborated curriculum content specifications and the role of teachers and curriculum specialists in selecting items. Although the greater homogeneity of the early grades (as contrasted to high school) curriculum in mathematics and reading (though not necessarily in social studies or other areas) may make this task easier for ECLS to accomplish, producing tests that are thoroughly curriculum-relevant and that truly measure school achievement will remain a pivotal concern.

References

Hamilton, L.S., Nussbaum, M., Kupermintz, H., Ennis, M.M., Kerkhoven, J.I.M., & Snow, R.E. (1994). *Enhancing the validity and usefulness of large-scale educational assessments: II. NELS:88 science achievement*. Stanford, CA: Stanford University Center for Research on the Context of Secondary Teaching.

Henderson, A., Abbott, C., & Strang, W. (1993). *Summary of the bilingual education state educational agency program survey of states' limited English proficient persons and available educational services 1991-92*. Report to the Office of Bilingual Education and Minority Languages Affairs, U.S. ED submitted by Development Associates, Inc. (Westat, Inc., subcontractor).

Ingels, S.J. (1991). *The problem of baseline excluded students in a school-based longitudinal study*. Paper presented at the American Educational Research Association annual meeting. ERIC ED 334-273.

Ingels, S.J. (1996). *Sample exclusion in NELS:88: Characteristics of base year ineligible students; changes in eligibility status after four years*. Washington, D.C.: National Center for Education Statistics. NCES 96-723.

Ingels, S.J., et al. (1987). *NELS:88 base year field test report*. Chicago, IL: NORC. See particularly Rock, D.A. & Pollack, J.M., "The Cognitive Test Battery."

Ingels, S.J., Scott, L.A., Rock, D.A., Pollack, J.M. & Rasinski, K.A. (1994). *NELS:88 base year to first follow-up final technical report*. Washington, D.C.: National Center for Education Statistics.

Ingels, S.J., Scott, L.A., and Frankel, M.R. (1996). *NELS:88 second follow-up: Sampling design, weighting and estimation report*. Chicago: NORC report to the National Center for Education Statistics.

Ingels, S.J., & Owings, J.A. (1995). Methodological Issues Encountered in Following a Cohort of Eighth Graders. In *Statistical policy working paper 23: Seminar on new directions in statistical methodology*. Washington, D.C.: Statistical Policy Office, U.S.Office of Management and Budget.

Kupermintz, H., Ennis, M.M., Hamilton, L.S., Talbert, J.E., & Snow, R.E. (1994). *Enhancing the validity and usefulness of large-scale educational assessments: I. NELS:88 mathematics achievement*. Stanford, CA: Stanford University Center for Research on the Context of Secondary Teaching.

McGrew, K.S., Thurlow, M.L., & Spiegel, A.N. (1993). An investigation of the exclusion of students with disabilities in national data collection programs. *Educational Evaluation and Policy Analysis, 15,*339-352.

Mullis, I.V.S., Dossey, J.A., Owen, E.H., & Phillips, G.W. (1993). *NAEP 1992 mathematics report card for the nation and the states.* Washington, D.C.: National Center for Education Statistics.

Muraki, E.J. & Bock, R.D. (1991). *PARSCALE: Parameter scaling of rating data.* Chicago, IL: Scientific Software, Inc.

Pollack, J.M., and Rock, D.A. (Forthcoming, 1997). *Free response tests in NELS:88.* Washington, D.C.: National Center for Education Statistics.

Qian, J. (1996). A Comparison of Weights Derived From Different Models. *1996 Proceedings of the Survey Research Methods Section of the American Statistical Association.*

Rock, D.A. & Pollack, J.M. (1991). *NELS:88 base year psychometric report.* Washington, D.C.: National Center for Education Statistics. NCES 91-468.

Rock, D.A. & Pollack, J.M. (1995). *NELS:88 base year to second follow-up psychometric report.* Washington, D.C.: National Center for Education Statistics. NCES 95-382.

Spencer, B.D. & Foran, W. (1991). Sampling probabilities for aggregations, with application to NELS:88 and other education longitudinal surveys. *Journal of Educational Statistics, 16* (1), 21-33.

Spencer, B.D., Frankel, M.R., Ingels, S.J., Rasinski, K.A. & Tourangeau, R. (1990). *NELS:88 base year sample design report.* Washington, D.C.: National Center for Education Statistics.

Thurlow, M.L., Ysseldyke, J.E., & Silverstein, B. (1993). *Testing accommodations for students with disabilities: A review of the literature.* Minneapolis, MN: National Center on Educational Outcomes.

Willingham, W.W., Ragosta, M., Bennett, R.E., Braun, H., Rock, D.A., & Powers, D.E. (1988). *Testing handicapped people.* Needham Heights., MA: Allyn and Bacon.

Ysseldyke, J.E. & Thurlow, M., eds. (1993). *Views on inclusion and testing accommodations for students with disabilities.* Minneapolis, MN: National Center on Educational Outcomes.

2.7 The Canadian National Longitudinal Survey of Children

Purpose of the Study

Canada's National Longitudinal Survey of Children (NLSC) is a new survey that was fielded for the first time in the fall of 1994. The survey was developed by Human Resources Development Canada and Statistics Canada under the auspices of the "What Works for Children—Information Development Program" of "Brighter Futures," a series of government initiatives designed to improve the health and well-being of Canada's children. The NLSC surveyed approximately 25,000 children, ranging in age from newborn to 11 years. Following the first wave of data collection, the NLSC will be repeated at two-year intervals to follow the children surveyed in 1994-95 into adulthood. Over time, the sample will be supplemented to provide cross-sectional as well as longitudinal data.

The survey gathers information for policy and program development on critical factors the affect the development of Canadian children. The survey covers a broad range of characteristics and factors that affect children's growth and development, such as children's families, neighborhoods, and schools, as well as children's health, temperament, behavior, child care and school experiences, participation in activities, and family and custody history. The primary objective of the NLSC is to develop a national database on the characteristics and life experiences of Canadian children as they grow from birth to adulthood. More specifically, the survey attempts to determine the prevalence of various biological, social, and economic risk factors among Canadian children and youth, and to monitor their impact on children's development.

The survey provides national and, as far as possible, provincial-level data. Children from the Yukon and Northwest Territories as well as the Canadian provinces are surveyed. Options for developing a separate longitudinal survey of Indian and Inuit children who are currently living on reservations is being investigated since these populations are not covered by the sample selection method being used for the NLSC; the feasibility of extending the survey to off-reserve aboriginal populations is also being considered. Supplements to the NLSC are also under discussion; one would examine intergenerational literacy, while a second would study an augmented representative sample of recent immigrant children. Provincial buy-ins for additional augmented samples are also being considered.

Sample Design

The 1994 sample of the NLSC includes approximately 25,000 children between the ages of 0 and 11 years. Information was collected on up to four children per household for selected households throughout Canada; in households with more than four children under the age of 12, four children were randomly selected. Participating households were selected from Statistics Canada's Labour Force Survey sample frame. The sample is divided into seven age groupings: 1, 2-3, 4-5, 6-7, 8-9, and 10-11. The children in the original sample are to be surveyed at two-year intervals until adulthood. The sample will be augmented for age groups no longer covered by the longitudinal survey to maintain coverage of the lower age ranges for cross-sectional

purposes. Children added to the sample for cross-sectional purposes will not be followed longitudinally. Replacements for the original respondents will be considered if attrition rates are concentrated within particular populations. Respondents who move between cycles will be tracked, and tracking will be facilitated by obtaining the names of two contact persons and their telephone numbers during each wave of data collection.

The sample and content of the NLSC have been partially integrated with the National Population Health Survey (NPHS), first implemented in June 1994, to allow both surveys to collect information regarding children's health while minimizing respondent burden. Respondents for the NPHS are contacted every two years. Four data collection periods per cycle are planned for NPHS. Data collection for the first cycle occurred in June 1994, August 1994, November 1994, and March 1995. For November 1994 and March 1995, data was collected for both NLSC and NPHS. A total of 3,000 NLSC households had a child chosen as the selected respondent for the NPHS. The NLSC survey instruments were administered for that child and for all other children in the household within the designated age range, to a maximum of four children per household. Data on approximately 5,000 children were thus collected for the NLSC through the integrated collection. Common areas of interest to the two surveys are covered by a standard set of survey questions. Data collection for the first cycle of the NLSC occurred in November 1994 (first integrated collection), December and February (main collection), and March 1995 (second integrated collection).

Assessment Instruments and Procedures

The instruments selected for the NLSC are designed to measure characteristics of children's families, neighborhoods, and schools, as well as characteristics specific to the child, and emphasize children's socioemotional and physiological well-being. For 4 and 5 year olds, children's receptive vocabulary skills are also measured, and for children who are in school, information on children's academic performance is obtained. For the most part, the survey is conducted in children's households and the parent most knowledgeable about the child is asked to respond on the child's behalf. The Peabody Picture Vocabulary Test is directly administered to children who are 4 to 5 years of age. For 10 and 11 year olds, a self-completed questionnaire is also used. The feasibility of teacher and principal self-completed questionnaires was examined.

Selection criteria. The following criteria were used to delineate the themes to be addressed by the survey instruments and to set priorities for content selection: (1) the particular concept to be explored should address an important policy or scientific issue; (2) the content addressed should cover risk factors, protective factors, and child outcomes; (3) the concepts covered should concern a significant segment of the population; (4) the data required to address particular concepts should be easily obtained within the context of a household survey. The selection of instruments for the NLSC was also guided by several criteria: (1) conciseness of the measures; (2) suitability for use in a household survey; (3) suitability for use by lay interviewers with a cross-section of the Canadian population (i.e., with children from various ethnocultural and socioeconomic backgrounds); (4) comparability with measures used in other studies conducted in

Canada and abroad; and (5) appropriateness for both longitudinal and cross-sectional use (i.e., chosen measures are applicable through each child's development as well as comparable across different groups at one point in time). Other requirements for the selection of instruments included complete documentation regarding the psychometric properties of selected scales, testing modified or adapted measures to ensure that reliability and validity of the original measures was maintained, and availability of the instruments in both French and English (the official languages of Canada).

Instruments and procedures. The NLSC consists of seven computer-assisted personal interviews and four self-administered questionnaires. Personal interviews are used in order to secure the participation of households over a long period of time as well as to develop a rapport between respondents and the interviewers. The nature of some elements of the NLSC also make it necessary for data collection to occur through personal interviewing. For example, the PPVT are individually administered by interviewers to 4 and 5 year olds. The various instruments that are used in the NLSC are described below.

Household record. The household record is used in all NLSC and NPHS collections to obtain information on the age, sex, and marital status of all household members. Since the households selected for NLSC are households that have participated in Statistics Canada's Labour Force Survey, this information will already have been collected for each household. Interviewers simply verify the information and revise it as required. Additional information is collected on relationships among household members, rental or ownership of houses or apartments, and size (number of bedrooms) and condition of family dwellings. These questions obtain information on housing conditions that have an impact on children's well-being.

General questionnaire. The general questionnaire is also an integrated NLSC/NPHS questionnaire that is used to gather information on specific elements of the child's environment. For the NLSC, the General Questionnaire collects information on sociodemographics (e.g., country of birth, year of immigration, ethnicity, religious affiliation, primary language spoken, and other languages spoken by household members); certain family characteristics such as parents' education, labor force activity, and sources and amounts of household income; and two areas of adult health: restriction of activities and chronic conditions. Other aspects of the family and neighborhood are covered in the Parent Questionnaire (see below).

Parent questionnaire. The parent questionnaire provides additional information on the child's environment, including parents' physical and mental health, family functioning, presence of social supports, and characteristics of the neighborhood. Parent health questions address the general state of health of both the respondent and his or her spouse/partner. Individual histories of cigarette smoking and alcohol consumption are also collected. For mothers of children under two years of age,

a pregnancy history is also obtained. Questions on general health, smoking, and alcohol consumption are from the National Population Health Survey; questions about pregnancies and births were developed by Dr. J.F. Saucier of St. Justine Hospital, Montreal.

Because it was decided that it would be most appropriate to measure one particular aspect of mental health rather than attempting a global measure of mental well-being, health questions focus on symptoms of depression exhibited by the respondent because of their prevalence and established impact on children. A shortened version of the Centre for Epidemiological Study Depression Scale, developed by L.S. Radloff, is used to measure symptoms associated with depression within the previous week. Questions regarding family functioning address problem solving, communication, roles, affective responsiveness, affective involvement, and behavior control. These questions were developed by researchers at Chedoke-McMaster Hospital, McMaster University, and have been widely used both within Canada and abroad. Several questions address respondents' satisfaction with their neighborhoods as places to raise children, and cover length of residency in the neighborhood, safety, social cohesion, and neighborhood problems. The questions represent a revised version of specific sections of the Smicha-Fagan Neighborhood Questionnaire used by Dr. Jacqueline McGuire in her studies of neighborhoods in Boston and Chicago. Revisions are based on a factor analysis of the specific sections and were made in consultation with Dr. McGuire.

A shortened version of the Social Provisions Scale, developed by Drs. Carolyn Cutrona and Daniel Russell of Iowa State University, is used to measure perceived social support. The shortened version of the scale was developed for the Ontario Better Beginnings, Better Futures Project and focuses on the following aspects of social relationships: guidance, attachment, and reliable alliance (the assurance that others can be counted on for practical help). In most cases questions from other surveys have been added to the original scales and questionnaires mentioned above.

Children's questionnaire. The children's questionnaire collects information from the person most knowledgeable about the child on a broad range of child characteristics and contextual factors, including children's health, temperament, behavior, literacy environment, education, involvement in nonschool activities, social relationships, child care experiences, family and custody history, and parenting styles and behaviors. Questions regarding children's physical health cover general health, injuries, limitations and chronic conditions, and use of health services and medications. For children four years of age and older, information is also collected on hearing, sight, speech, and overall mental well-being. For children under three, data are collected on factors such as length of gestation and weight at birth. For children under two, information is also collected on delivery, general health of the child at birth, and specialized services following the birth. The Infant Characteristics Questionnaire, developed by John Bates of Indiana

University, is used to measure the temperament of children under four years of age. A revised version of the scale, developed by Dr. Jo-Anne Finegan at Toronto's Hospital for Sick Children, is used for three year olds.

Information on children's educational experience varies with the age of the child; more information is collected for older children who have greater school experience. Basic information is collected on children's grade level, type of school and language of instruction, behavior problems at school, absences, parents' educational aspirations for their children, and number of school changes and residential moves. For children who have begun formal schooling, additional questions address skipped or repeated grades, achievement, special education, parents' perceptions of the school climate, and the importance of good grades to parents. Additional information on children's school achievement and behavior is obtained from teachers. Measures of children's literacy environment include children's exposure to books, their interest in reading, parental encouragement of children's writing skills, and the frequency with which children are given homework assignments. Questions regarding children's out-of-school activities include children's participation in organized group activities, TV viewing habits, household responsibilities, amount of time spent in playing alone and with friends, and summer activities.

Several measures are used to assess children's behavior. For children younger than four, questions focus on sleep and eating patterns. For children over the age of two, the frequency of other specific behaviors, as noted by the parents, is collected. Information is also collected by self-report for 10- and 11-year-olds. The following behaviors are measured for 4- to 11-year-olds: conduct disorder, hyperactivity, emotional disorder, anxiety, indirect aggression, physical aggression, inattention, and prosocial behaviors. For two and three year olds, separation anxiety and opposition are added to the measures already mentioned; indirect aggression and some aspects of conduct disorder are not measured. Parents of 10- and 11-years-olds are asked additional questions regarding their children's behavior; these questions parallel those asked in the self-completed questionnaire for 10- and 11-year-olds (see below).

Questions regarding children's social relationships focus on how the child gets along with parents, siblings, teachers, friends, and classmates. Information on other important adults in the child's life is also obtained. Parents' knowledge of the names of friends of 8-9- and 10-11-year-olds is also investigated, along with the parents' perceptions of these other children's behavior, and whether their own child is shy or outgoing.

Questions regarding child care focus on the types provided to children while parents are working or studying, and the child's past experiences with child care. The amount of time spent by the children in child care and the methods of care

used for each child are assessed. Information is also obtained on the number of changes in child care arrangements that the child has experienced and the reason(s) for changes that have occurred in the past 12 months. Questions regarding children's family and custody history address significant family restructuring events such as marital separation, divorce, and remarriage that have occurred before or after the child entered the family. Only the parent in the selected household is interviewed; in cases where parents are divorced but have joint custody, the other parent is not interviewed. Measures of parenting behaviors focus on positive interactions, consistent parenting and hostile or ineffective parenting, and aversive and nonaversive parenting techniques. Several sources were used in developing questions for the children's questionnaire, among them, the Canadian Survey of Labour and Income Dynamics, the U.S. National Assessment of Educational Progress, and the NLSY79 Child Assessments.

The Vineland Adaptive Behavior Scales. The Vineland Adaptive Behavior Scales were developed by Sara Sparrow, David Balla, and Domenic Cicchetti at Yale University. The scales measure aspects of children's social and physical development and can be used from birth to adulthood. For NLSC, the person most knowledgeable about the child is asked to complete the scales for children under four. The scales assess behavior within four domains: (1) "communication" deals with how the child speaks and understands others; (2) "daily living skills" deals with practical skills needed to take care of oneself; (3) "socialization" deals with skills needed to get along with others, play activities, and use of leisure time; (4) "motor skills" assesses children physical skills and motor development.

The Peabody Picture Vocabulary Test (PPVT). The PPVT is designed to measure children's receptive or hearing vocabulary and indicates the extent of the child's language acquisition. The PPVT may be used with any age group; for NLSC, it is administered to four and five year olds. A French adaptation of the PPVT has been developed; both French and English versions are used in the NLSC.

Administrative information. The following administrative information is collected after interviews have been completed with parents: (1) respondents' permission to share data with Human Resources Development Canada; (2) name, address, and telephone number of two persons who know the respondent and who can be contacted in the event that the respondent moves prior to the next wave of data collection; (3) respondent's consent that the child's teacher may be contacted to complete a questionnaire; and (4) an indication by the interviewer of whether first contact was made by telephone or in person.

Interviewer questionnaire. This questionnaire consists of a set of questions concerning the interviewer's observations of the respondent's neighborhood, which are from the Neighborhood Cluster Observation Schedule used by Dr. Jacqueline

McGuire in her studies of neighborhoods. Interviewers are asked to assess factors such as traffic volume, presence of garbage and needles/syringes on sidewalks, visible signs of loitering, visibly threatening or drunken behavior, land use on the block or road, and the condition of buildings. The information in the interviewer questionnaire supplements information on the neighborhood provided by parents.

Self-completed questionnaire for 10-11-year-olds. The self-completed questionnaire for children is designed to collect information directly from older children that supplements information obtained from parents and teachers. The questionnaire also collects unique information from the child for topics on which only the child can reliably report. Parents must give consent for their children to complete the questionnaire. Written instructions are included and children are encouraged to complete the questionnaires in a private setting. Completed questionnaires are sealed in an envelope to ensure confidentiality.

The topics covered by the questionnaire include children's relationships with family members and friends (e.g., number of friends, time spent with friends, and the quality of the child's relationships with parents, peers, and teachers); children's attitudes toward school, their perceptions of how they are doing in school, and their feelings of safety and social acceptance; children's perceptions of the teacher with respect to fairness and providing extra help; children's perceptions of support provided by parents for school-related work including help, encouragement, and performance expectations; the consistency with which children complete homework assignments, and the availability of a place at home to do homework.

Other questions replicate items used in the other NLSC questionnaires. Children are asked to complete a behavioral checklist that is also included in the Children's Questionnaire and the Teacher Questionnaire. Questions addressing the child's relationship to his or her parents complement questions asked in the parenting section of the parent-completed Children's Questionnaire. Other questions address the children's use of cigarettes, alcohol, and drugs, and their frequency of use by both the child and his or her friends; the nature and extent of children's participation in extracurricular activities (e.g., sports, music, Guides or Scouts); and children's overall sense of self-esteem and their perceptions of their physical appearance. Key physiological indicators of puberty are also included as questionnaire items. Several of the questions included in the Children's Self-Completed Questionnaire are from the Marsh Self-Description Questionnaire and the World Health Organization Survey of Health Behaviors in School Children. Questions from several other surveys have also been used.

Teacher's questionnaire. The teacher's questionnaire is designed to measure the academic achievement and behavior of school-age children as a cross-reference to parents' perceptions. This information also complements information obtained in the self-completed questionnaires for 10-11-year-olds. The teacher questionnaire

is mailed to the teacher of every school-aged child in the survey whose parents have given consent. A wide variety of information is collected on the child's educational development, including grade level, skipped or repeated grades, academic performance, time spent on various subjects, language of instruction, personal/social skills, work habits, special skills and talents, enhanced instruction, and special education. The questionnaire also covers parental involvement with the school, characteristics of the class, the teacher's instructional practices, the teacher's perceptions of the school, and selected demographic characteristics of the teacher (age, gender, education, and teaching experience). Many of questions asked were developed specifically for the NLSC; sources for other questions include the Ontario Child Health Survey (special education questions) and the Ontario Tri-Ministry Project teacher questionnaire.

Principal's questionnaire. The purpose of the principal's questionnaire is to gather information on the school environment. Consequently, the questionnaire focuses on school policies and educational climate rather than on specific characteristics of the child. Topics covered include general information on the students (e.g., languages spoken by students, family backgrounds, student disabilities), characteristics of the school (e.g., school enrollment, methods of assigning students to classes, rates of absenteeism, the extent and nature of disciplinary problems at the school), the principal's perceptions of the school, and levels of parental support including volunteering for school activities and the strength of the parent-teacher association. Demographic information about the principal is also collected. Once parents' consent is obtained, questionnaires are mailed to the principals of those schools attended by one or more children in the NLSC sample. Many of the questions were developed specifically for NLSC; sources for other questions include the Third International Mathematics and Science Survey and Dr. Douglas Willms's Principal Survey. Dr. Willms is a member of the Expert Advisory Group of the NLSC and is affiliated with the Centre for Policy Studies in Education at the University of British Columbia.

Publications and Uses of the Data

An overview of survey instruments (1994a) used in the NLSC and copies of the questionnaires (1994b) used in the July 1994 field test are available from the Department of Human Resources Development and Statistics Canada. Final revisions to the survey instruments were to be made based on the results of the July field test and were scheduled to be completed in late August or early September of 1994. Several other tests of survey instruments have been conducted. Draft questionnaires were tested through personal interviewing and focus group discussions in Toronto, Peterborough, and Montreal in June and August of 1993. The tests provided initial feedback on the sensitivity of content, respondent understanding of wording, and general reactions to the survey instruments, and revisions to the instruments were made based on the feedback. A preliminary field test was carried out in November 1993 in 150 households in Winnipeg and Toronto/Hamilton. Several sets of focus tests were also conducted in May and June

1994. The teacher and principal questionnaires were tested at sites in Ontario and Quebec. The self-completed questionnaire for 10- and 11-year-olds was tested in Montreal, Ottawa, Toronto, and Peterborough. In June, the complete set of survey instruments was tested in Montreal, Toronto, and Halifax, and further revisions were made to the survey instruments based on the results of these tests. Details regarding the July 1994 field test are not yet available.

Implications for ECLS

The NLSC parent, teacher, and principal questionnaires will be useful in developing interview questions for ECLS. Detailed summaries of the objectives, selection criteria, and sources of information for each question used are included in Statistics Canada's overview of survey instruments (1994a) and will be useful in guiding the selection of questions for ECLS. Because the two studies are similar in size, guidelines for balancing interview time, cost, collection methodology, and the use of supplemental samples may also prove useful for ECLS. Apart from the PPVT, no direct assessments of children's cognitive skills are used in the NLSC.

References

Department of Human Resources Development and Statistics Canada. (1994a). *National Longitudinal Survey of Children: Overview of survey instruments.* Catalog Number 94-02 (June 1994). Ottawa, Ontario: Government of Canada.

Department of Human Resources Development and Statistics Canada. (1994b). *National Longitudinal Survey of Children: Survey instruments for July 1994 field test.* Catalog Number 94-01 (June 1994). Ottawa, Ontario: Government of Canada.

2.8 The National Survey of Children

Purpose of the Study

The National Survey of Children (NSC) was designed to provide a broad profile of the physical health, emotional well-being, social development, and academic achievement of elementary school children in the United States, as well as to assess the family and neighborhood circumstances in which these children were growing up. The first wave of interviews was conducted in 1976 with children from 7 to 11 years of age. The parent most able to provide information about each child was also interviewed, and the children's teachers completed self-administered questionnaires. The second wave of the survey was fielded in 1981 when the children were 12 to 16 years of age and focused on the effects of marital disruption and changing family structure on adolescents' behavior and emotional well-being. The third wave was conducted in 1987 when respondents were between 18 and 22 and focused on welfare dependence, the social development and well-being of young adults, and their early sexual and fertility behavior.

Funding for the first wave of National Survey of Children was provided by the Foundation for Child Development. The second wave of the survey was jointly sponsored by the Foundation for Child Development and the National Institute of Mental Health. Funding for the third wave was provided by the National Institute of Child Health and Human Development, the Department of Health and Human Services, the Robert Wood Johnson Foundation, and the Ford Foundation. Data collection for all three waves was conducted by the Institute for Survey Research at Temple University.

Sample Design

The original 1976 sample was a multi-stage stratified probability sample of households containing children aged 7 to 11 (i.e., born between 1965 and 1970). Up to two children per household were eligible to be interviewed; if a selected family had three or more eligible children, two were randomly selected for the study. As a result of these procedures, 2,193 households were located, and interviews were completed with 2,301 children from 1,747 households resulting in a completion rate of 80 percent. Male and female children were equally represented. Black households were intentionally oversampled resulting in the inclusion of approximately 500 black children in the original sample. No oversampling was done for families of Hispanic or Asian origin. The data were weighted to correct for the oversampling of black households and for other minor differences between sample and census estimates by age, sex, and place of residence. Personal interviews were conducted with the children themselves and with the parent most knowledgeable about the child (usually the mother). A follow-up study of schools attended by the children was conducted in 1977. School information, obtained from the child's teacher, was collected for 1,682 children (74 percent of the sample).

The second wave of the study was completed in 1981. Due to funding limitations, 1,749 children were selected for restudy, and telephone rather than in-person interviews were conducted. Data were again collected from the child, a parent, and a teacher. All children from high-conflict

BEST COPY AVAILABLE

or disrupted families were followed and a subsample of other children were selected for restudy. Interviews were conducted with 1,423 children, or 79 percent of the selected subset. The data were weighted to adjust for differential subsampling and completion rates.

A third wave of the study was completed in 1987 with 1,147 youth respondents, 80 percent of those eligible to be reinterviewed. Telephone interviews were conducted with the young adult and with the most knowledgeable parent or guardian. The data for the third wave were weighted to adjust for differential attrition as well as the oversampling of black children, the undersampling of children from large families, and the oversampling of children from high-conflict and disrupted families in Wave II. Overall attrition from the initial completed set of cases to the Wave III interviews was 36 percent. Attrition rates for black families and those who were informally separated or never married were higher than those for white families and families where parents remained married or were formally divorced. No additional data collection is planned.

Assessment Instruments and Procedures

Data collection for the initial round of the National Survey of Children was carried out through personal interviews with children and their parents, and through self-administered questionnaires with teachers. Telephone interviews were conducted with respondents in subsequent waves. Because the original survey focused on elementary school children between the ages of 7 and 11, and is most relevant for ECLS, this summary will focus on the topics covered in that survey.

Parent interviews. The 1976 parent interviews provide a particularly rich source of data on the parents and families of children in the study. Information was obtained on the national origins of the parents' ancestors, the religions in which the parents were raised, the types of places they grew up, their patterns of residential mobility, and their educational attainments and occupations. Detailed marital and parenting histories were also collected. Information was obtained not only for parents living in the household but for those living elsewhere because of divorce or separation. Interviewers asked about the respondent's functioning and well-being (usually the mother), and specifically inquired about financial worries, time pressures, physical health, feelings of anxiety, depression, and exhaustion, and overall life satisfaction. The parent was also asked to report on neighborhood characteristics such as crime, noise, and dangers to children, and on the quality of elementary schools and public services. Questions about the family focused on family activities, childcare arrangements, parents' childrearing goals and educational aspirations for their children, and areas of marital conflict. Parents were also asked to provide a history of their child's injuries and accidents, and to detail any mental, physical, or emotional limitations or conditions that might interfere with the child's schoolwork or prevent the child from participating in play activities. Parents also assessed the amount of time their children spent on homework and various leisure activities, their children's problem behaviors, their academic performance and progress in school, their needs for special classes or assistance, and their opportunities to participate in classes or activities outside of school. Questions regarding the quality of parent-child relationships focused on time spent playing with children, teaching them new skills, and helping them with homework; familiarity with the child's friends; typical reactions

to misbehavior; attention to the child's physical and emotional needs, including medical and dental care, parental supervision, and special counseling or treatment; and feeling close to the child.

Child interviews. The Wave I interviews with children addressed a similar range of topics. Questions regarding the family focused on rules at home, typical responses by each parent to good and bad behavior, responsibilities around the house, and the quality of the child's relationship with family members (e.g., time spent together, help given with homework, and the frequency of arguments or fights). With regard to the neighborhood, children were asked if there were other children they could play with, whether there were adults outside the family children liked to talk to and could spend time with, whether they had ever been bothered or threatened by other children or adults, and whether there was something they would change about the neighborhood to make it a nicer place for children. Several questions addressed children's friendships and peer relationships. Children were asked if they usually played alone or with friends, whether they had a best friend or several friends they liked equally, whether they played with children of the same age or with younger or older children, and whether they often played with children of the opposite sex or with children of a different race. Interviewers also asked children how often they argued or fought with friends. With regard to school, children were asked whether they were interested in their schoolwork, whether they liked and got along with their classmates and teachers, and whether they and their classmates usually paid attention in class. Interviewers also asked children to rate their academic performance (i.e., one of the best students in the class, above or below the middle, near the bottom) and their skills in various academic subjects. A number of questions also addressed children's feelings about themselves and their fears and worries about schoolwork, friends, and family.

Teacher questionnaires. The child's teacher (or main teacher if there was more than one) was asked to report on the child's academic performance (e.g., to provide grades and standardized test scores, to note class rank and promotions or retentions), and to detail absences, behavioral problems and disciplinary actions, and special conditions, such as a physical handicap or learning disability that interfered with the child's schoolwork or limited his or her participation in play activities. Teachers were also asked to assess the need for, availability, and use of special resources for each child in the study (e.g., advanced or remedial instruction and special facilities for physically handicapped or learning disabled students). Background information was also obtained on the schools such as subjects taught, ability groupings, the marking system, and on the teachers themselves such as schools attended and degrees received, number of years as a full-time teacher, gender, ethnicity, relationship to the child (e.g., classroom teacher, special education teacher), and length of acquaintanceship with the child.

No information is available on the time required to complete the parent and child interview or the teacher questionnaires. The protocols themselves are quite lengthy, however, and would probably take an hour or more to complete.

Interviewer evaluations. Interviewers provided evaluations of the conditions under which parent and child interviews were conducted (e.g., the presence of family members or friends, and their comments or contributions), and noted the respondent's ethnic group, physical attractiveness,

obvious physical handicaps or exceptional physical characteristics, and apparent intelligence. They also assessed the respondent's attitude toward being interviewed, his or her attentiveness to the interviewer's questions, and the apparent truthfulness and sincerity of his or her answers. In addition, interviewers recorded their observations of the general atmosphere of the household and the presence of educational resources, such as reading materials and educational games and toys, and provided a description of the family dwelling (e.g., single family house, apartment building) and its general maintenance as well as a description of the street on which the dwelling was located.

Publications and Uses of the Data

The data from the National Survey of Children have not been widely used. Zill and Daly (1993) provide a list of 13 publications based on the NSC data; the majority of articles listed focus on the effects of marital disruption and parental conflict (see, e.g., Furstenberg, Nord, Peterson and Zill, 1983; Peterson and Zill, 1986). The data for all three waves are available for secondary analysis. Response frequencies for items from the initial wave of the study have been compiled and are available from Child Trends.

Implications for ECLS

Because of their comprehensiveness, the parent interviews and teacher questionnaires will undoubtedly be of use in developing interview questions for ECLS. The questions assessing children's problem behaviors were developed into the Behavior Problems Index used in the NLSY79 Child Assessments and, as noted in the NLSY79 study summary, might be used for ECLS. The parent and child interviews and teacher questionnaires are nicely articulated, making it possible to identify convergent or divergent views on children's academic performance, and their relationships to family members, peers, and teachers. Teachers, however, do not provide an assessment of parents' interest and involvement in their children's schooling. The study is of limited usefulness in other respects. The child interviews are lengthy and could not be conducted with younger children; the questions themselves are in many cases inappropriate for preschool or kindergarten children; and the information obtained is somewhat unreliable. As Zill and Daly (1993) observe, children's responses are more mercurial than those of adults; consequently scales based on child interviews tend to be less reliable than those based on interviews with older respondents. They note, however, that "meaningful relationships between family characteristics and parenting variables based on children's reports have been found" (p. 288). Although information on children's academic performance was obtained from teachers (both grades and standardized test scores), no independent assessments were conducted for NSC.

References

Furstenberg, F.F., Jr., Nord, C.W., Peterson, J.L., & Zill, N. (1983). The life course of children of divorce: marital disruption and parental conflict. *American Sociological Review*, *48*, 656-666.

Peterson, J.L. & Zill, N. (1986). Marital disruption, parent-child relationships, and behavioral problems in children. *Journal of Marriage and the Family*, *48*, 295-307.

Zill, N. & Daly, M., eds. (1993). *Researching the family: A guide to survey and statistical data on U.S. Families*. Washington, D.C.: Child Trends.

2.9 National Child Development Study

Purpose of the Study

Data collected in the 1991-92 round of the National Child Development Study provide a unique resource because the original respondents were selected as infants in 1958, and contacted in a series of followups, so in 1991 most respondents were age 33. During this round of data collection, the study included the children of the original respondents, and conducted assessments that parallel those included in NLSY79. Thus, the British study provides an international comparison group for NLSY79. But even more impressive, NCDS5 (the age 33 fifth follow-up) provides researchers with data from direct, early cognitive assessments for two generations—parents and children.

History. The National Child Development Study (NCDS) has its origins in the Perinatal Mortality Survey (PMS). Sponsored by the National Birthday Trust Fund, the study was designed to examine the social and obstetric factors associated with stillbirth and death in early infancy among the 17,000 children born in Great Britain born during the week of March 3 to March 9, 1958. It was the second in a series of three such perinatal studies, the others being based on a week's births in 1946 and 1970. Each has formed the basis of a continuing longitudinal study (Shepard, 1985).

In 1964, the Department of Education and Science agreed to commission the National Children's Bureau to collect information on all these children when they were seven. The study then became known as the National Child Development Study. In 1985 the Social Statistics Research Unit (SSRU) of City University took over day-to-day responsibility for the NCDS (NCDS News, Autumn 1987).

In all, there have been five attempts (at ages 7, 11, 16, 23, and 33) to trace all members of the original study in order to monitor their physical, educational, and social development. In addition, in 1978 (age 20), contact was made with the schools attended by members of the birth cohort at the time of the second follow-up in 1974 (age 16) in order to obtain details of public examination entry and performance. Similar details were also sought from colleges where these were identified by schools (Shepard, 1985).

Sample Design

NCDS is a longitudinal study which takes as its subjects all those living in Great Britain who were born between March 3 and 9, 1958. The sample for the first three follow-ups also included immigrants to Great Britain who were born during the sample time period. The sample for the fourth and fifth follow-ups differed from prior surveys in that it consisted of all those who had participated in at least one of the earlier NCDS follow-ups, excluding those subjects known to have emigrated or to have died. There was no attempt to include new immigrants, as there had been with the first three follow-ups.

Assessment Instruments and Procedures

In each of the first three follow-ups, information was obtained from four main sources: the subjects themselves, the parents, local authority medical officers, and schools/teachers. The children (subjects) were given tests of attainment which included:

Age seven:

Southgate Reading Test—a test of word recognition and comprehension.

Copying Designs Test—to obtain some assessment of the child's perceptual-motor ability.

Drawing A Man Test—as an indication of the child's general mental and perceptual ability.

Problem Arithmetic Test—a general test of arithmetic skills.

Age eleven:

General Ability Test—containing verbal and nonverbal items.

Reading Comprehension Test—constructed by the National Foundation for Educational Research in England and Wales (NFER) specifically for this study.

Arithmetic/Mathematics Test—again, constructed by the National Foundation for Educational Research in England and Wales (NFER) specifically for this study.

Age sixteen:

Reading Comprehension Test—same test used at age 11.

Mathematics Test—devised at the University of Manchester and originally intended for use in the NFER's study of comprehensive schools.

At ages 11 and 16 questionnaires were also administered to the subjects. At age 11 the questionnaire contained questions on leisure activities and attitudes toward school. Each child was also asked to write a short composition on the life he or she imagined for himself or herself at age 25.

At age 16 a more substantial questionnaire included questions on: attitudes about school and methods of punishment in school, future educational and occupational expectations and aspirations, reasons for leaving school and choosing a job, school absences, self-ratings in school

subjects, spare-time work, income and pocket money, intentions about marriage and having children, sex education and preparation for parenthood, leisure activities, family relationships, smoking and drinking, and handedness.

NCDS4. NCDS4 (age 23) differed from the earlier NCDS follow-ups in that information was obtained from the 1971 and 1981 Census as well as the subjects themselves.

NCDS5. NCDS5 (age 33) received funding from the National Institute of Child Health and Development (NICHD) in the U.S. to include a child supplement for the children of selected NCDS sample members. This funding was an extension of the support NICHD provides for the National Longitudinal Survey of Youth (NLSY79), which since 1979, has carried out annual surveys of a cohort in the age range 14-21. In 1986, 1988, 1990, 1992, 1994 and 1996, data collection included developmental assessments of the biological children of female respondents. The NICHD contribution included these measures in the NCDS5 to provide for collaborative and comparative work involving NCDS and the NLSY79. See the Study Summary for the National Longitudinal Survey Youth Cohort—Child Assessments for a complete detailed description of the child assessments. For NCDS5 the assessments were Anglicized versions of the NLSY79 assessments, that is, references to measurement and money were changed to the metric system and pounds from British Imperial and dollars in the mathematics subtest of the Peabody Individual Achievement Test. In the PPVT, four typically British words were substituted for their American counterparts (e.g., jug for pitcher).

The NCDS5 child sample differed from the NLSY79 child sample in that NCDS5 covered all children, natural and adopted, currently living with one-third of the sample (male and female) of cohort members. The NLSY79 child assessments are administered only to the biological children currently living with female respondents. The children of male respondents and adopted, step, or foster children are not included in the NLSY79 child sample. Information about the NCDS5 child was obtained from the child's mother even if the mother was not the cohort member, but the spouse or partner of the cohort member. Field work for NCDS5 was conducted during 1991 and early 1992. Data preparation was carried out in 1992 and 1993.

Publications and Uses of the Data

As part of the NCDS5 program funded by NICHD, the NCDS5 team will be providing an NCDS5 Mother and Child dataset which can be analyzed alongside that produced for the NLSY79 by the Center for Human Resource Research (CHRR) at Ohio State University. To date this dataset not been available to CHRR and there has been no analysis of the data.

Because the child supplement to the NCDS5 was modeled on the NLSY79 Child Study, and uses its the same assessments, methodological implications for ECLS are essentially similar to those of the NLSY79. However, as a rich source of cross-cultural comparison data, it will be of interest to take note of the research that this data set supports over coming years.

References

National Child Development Study Newsletter, Number 2, Autumn, 1987.

Shepard, P.M. (1985). *The National Child Development Study: An introduction to the background of the study and the methods of data collection.* Working Paper No. 1. London: Social Statistics Research Unit, City University.

Listing of NCES Working Papers to Date

Please contact Ruth R. Harris at (202) 219-1831
if you are interested in any of the following papers

Number	Title	Contact
94-01 (July)	Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association	Dan Kasprzyk
94-02 (July)	Generalized Variance Estimate for Schools and Staffing Survey (SASS)	Dan Kasprzyk
94-03 (July)	1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report	Dan Kasprzyk
94-04 (July)	The Accuracy of Teachers' Self-reports on their Postsecondary Education: Teacher Transcript Study, Schools and Staffing Survey	Dan Kasprzyk
94-05 (July)	Cost-of-Education Differentials Across the States	William Fowler
94-06 (July)	Six Papers on Teachers from the 1990-91 Schools and Staffing Survey and Other Related Surveys	Dan Kasprzyk
94-07 (Nov.)	Data Comparability and Public Policy: New Interest in Public Library Data Papers Presented at Meetings of the American Statistical Association	Carrol Kindel
95-01 (Jan.)	Schools and Staffing Survey: 1994 Papers Presented at the 1994 Meeting of the American Statistical Association	Dan Kasprzyk
95-02 (Jan.)	QED Estimates of the 1990-91 Schools and Staffing Survey: Deriving and Comparing QED School Estimates with CCD Estimates	Dan Kasprzyk
95-03 (Jan.)	Schools and Staffing Survey: 1990-91 SASS Cross-Questionnaire Analysis	Dan Kasprzyk
95-04 (Jan.)	National Education Longitudinal Study of 1988: Second Follow-up Questionnaire Content Areas and Research Issues	Jeffrey Owings
95-05 (Jan.)	National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors	Jeffrey Owings

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
95-06 (Jan.)	National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data	Jeffrey Owings
95-07 (Jan.)	National Education Longitudinal Study of 1988: Conducting Trend Analyses HS&B and NELS:88 Sophomore Cohort Dropouts	Jeffrey Owings
95-08 (Feb.)	CCD Adjustment to the 1990-91 SASS: A Comparison of Estimates	Dan Kasprzyk
95-09 (Feb.)	The Results of the 1993 Teacher List Validation Study (TLVS)	Dan Kasprzyk
95-10 (Feb.)	The Results of the 1991-92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation	Dan Kasprzyk
95-11 (Mar.)	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
95-12 (Mar.)	Rural Education Data User's Guide	Samuel Peng
95-13 (Mar.)	Assessing Students with Disabilities and Limited English Proficiency	James Houser
95-14 (Mar.)	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
95-15 (Apr.)	Classroom Instructional Processes: A Review of Existing Measurement Approaches and Their Applicability for the Teacher Follow-up Survey	Sharon Bobbitt
95-16 (Apr.)	Intersurvey Consistency in NCES Private School Surveys	Steven Kaufman
95-17 (May)	Estimates of Expenditures for Private K-12 Schools	Stephen Broughman
95-18 (Nov.)	An Agenda for Research on Teachers and Schools: Revisiting NCES' Schools and Staffing Survey	Dan Kasprzyk
96-01 (Jan.)	Methodological Issues in the Study of Teachers' Careers: Critical Features of a Truly Longitudinal Study	Dan Kasprzyk

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
96-02 (Feb.)	Schools and Staffing Survey (SASS): 1995 Selected papers presented at the 1995 Meeting of the American Statistical Association	Dan Kasprzyk
96-03 (Feb.)	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
96-04 (Feb.)	Census Mapping Project/School District Data Book	Tai Phan
96-05 (Feb.)	Cognitive Research on the Teacher Listing Form for the Schools and Staffing Survey	Dan Kasprzyk
96-06 (Mar.)	The Schools and Staffing Survey (SASS) for 1998-99: Design Recommendations to Inform Broad Education Policy	Dan Kasprzyk
96-07 (Mar.)	Should SASS Measure Instructional Processes and Teacher Effectiveness?	Dan Kasprzyk
96-08 (Apr.)	How Accurate are Teacher Judgments of Students' Academic Performance?	Jerry West
96-09 (Apr.)	Making Data Relevant for Policy Discussions: Redesigning the School Administrator Questionnaire for the 1998-99 SASS	Dan Kasprzyk
96-10 (Apr.)	1998-99 Schools and Staffing Survey: Issues Related to Survey Depth	Dan Kasprzyk
96-11 (June)	Towards an Organizational Database on America's Schools: A Proposal for the Future of SASS, with comments on School Reform, Governance, and Finance	Dan Kasprzyk
96-12 (June)	Predictors of Retention, Transfer, and Attrition of Special and General Education Teachers: Data from the 1989 Teacher Followup Survey	Dan Kasprzyk
96-13 (June)	Estimation of Response Bias in the NHES:95 Adult Education Survey	Steven Kaufman
96-14 (June)	The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component	Steven Kaufman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
96-15 (June)	Nested Structures: District-Level Data in the Schools and Staffing Survey	Dan Kasprzyk
96-16 (June)	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
96-17 (July)	National Postsecondary Student Aid Study: 1996 Field Test Methodology Report	Andrew G. Malizio
96-18 (Aug.)	Assessment of Social Competence, Adaptive Behaviors, and Approaches to Learning with Young Children	Jerry West
96-19 (Oct.)	Assessment and Analysis of School-Level Expenditures	William Fowler
96-20 (Oct.)	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96-21 (Oct.)	1993 National Household Education Survey (NHES:93) Questionnaires: Screener, School Readiness, and School Safety and Discipline	Kathryn Chandler
96-22 (Oct.)	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
96-23 (Oct.)	Linking Student Data to SASS: Why, When, How	Dan Kasprzyk
96-24 (Oct.)	National Assessments of Teacher Quality	Dan Kasprzyk
96-25 (Oct.)	Measures of Inservice Professional Development: Suggested Items for the 1998-1999 Schools and Staffing Survey	Dan Kasprzyk
96-26 (Nov.)	Improving the Coverage of Private Elementary-Secondary Schools	Steven Kaufman
96-27 (Nov.)	Intersurvey Consistency in NCES Private School Surveys for 1993-94	Steven Kaufman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
96-28 (Nov.)	Student Learning, Teaching Quality, and Professional Development: Theoretical Linkages, Current Measurement, and Recommendations for Future Data Collection	Mary Rollefson
96-29 (Nov.)	Undercoverage Bias in Estimates of Characteristics of Adults and 0- to 2-Year-Olds in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
96-30 (Dec.)	Comparison of Estimates from the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97-01 (Feb.)	Selected Papers on Education Surveys: Papers Presented at the 1996 Meeting of the American Statistical Association	Dan Kasprzyk
97-02 (Feb.)	Telephone Coverage Bias and Recorded Interviews in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-03 (Feb.)	1991 and 1995 National Household Education Survey Questionnaires: NHES:91 Screener, NHES:91 Adult Education, NHES:95 Basic Screener, and NHES:95 Adult Education	Kathryn Chandler
97-04 (Feb.)	Design, Data Collection, Monitoring, Interview Administration Time, and Data Editing in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-05 (Feb.)	Unit and Item Response, Weighting, and Imputation Procedures in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-06 (Feb.)	Unit and Item Response, Weighting, and Imputation Procedures in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97-07 (Mar.)	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97-08 (Mar.)	Design, Data Collection, Interview Timing, and Data Editing in the 1995 National Household Education Survey	Kathryn Chandler

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
97-09 (Apr.)	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
97-10 (Apr.)	Report of Cognitive Research on the Public and Private School Teacher Questionnaires for the Schools and Staffing Survey 1993-94 School Year	Dan Kasprzyk
97-11 (Apr.)	International Comparisons of Inservice Professional Development	Dan Kasprzyk
97-12 (Apr.)	Measuring School Reform: Recommendations for Future SASS Data Collection	Mary Rollefson
97-13 (Apr.)	Improving Data Quality in NCES: Database-to-Report Process	Susan Ahmed
97-14 (Apr.)	Optimal Choice of Periodicities for the Schools and Staffing Survey: Modeling and Analysis	Steven Kaufman
97-15 (May)	Customer Service Survey: Common Core of Data Coordinators	Lee Hoffman
97-16 (May)	International Education Expenditure Comparability Study: Final Report, Volume I	Shelley Burns
97-17 (May)	International Education Expenditure Comparability Study: Final Report, Volume II, Quantitative Analysis of Expenditure Comparability	Shelley Burns
97-18 (June)	Improving the Mail Return Rates of SASS Surveys: A Review of the Literature	Steven Kaufman
97-19 (June)	National Household Education Survey of 1995: Adult Education Course Coding Manual	Peter Stowe
97-20 (June)	National Household Education Survey of 1995: Adult Education Course Code Merge Files User's Guide	Peter Stowe
97-21 (June)	Statistics for Policymakers or Everything You Wanted to Know About Statistics But Thought You Could Never Understand	Susan Ahmed
97-22 (July)	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
97-23 (July)	Further Cognitive Research on the Schools and Staffing Survey (SASS) Teacher Listing Form	Dan Kasprzyk
97-24 (Aug.)	Formulating a Design for the ECLS: A Review of Longitudinal Studies	Jerry West

DEPARTMENT OF EDUCATION
UNITED STATES OF AMERICA